

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
20 December 2001 (20.12.2001)

PCT

(10) International Publication Number  
**WO 01/96400 A2**

(51) International Patent Classification<sup>7</sup>: **C07K 14/705**

(21) International Application Number: **PCT/US01/19464**

(22) International Filing Date: **18 June 2001 (18.06.2001)**

(25) Filing Language: **English**

(26) Publication Language: **English**

(30) Priority Data:  
60/212,331 16 June 2000 (16.06.2000) US  
60/269,758 16 February 2001 (16.02.2001) US

(71) Applicant (for all designated States except US): **MIL-LENNIUM PHARMACEUTICALS, INC.** [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **GLUCKSMANN, Maria, Alexandra** [AR/US]; 33 Summit Road, Lexington, MA 02173 (US). **SILOS-SANTIAGO, Inmaculada** [ES/US]; 18 Hillard Street, Cambridge, MA 02138 (US).

(74) Agents: **MANDRAGOURAS, Amy, E.**; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 et al. (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**

- without international search report and to be republished upon receipt of that report
- with (an) indication(s) in relation to deposited biological material furnished under Rule 13bis separately from the description

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **52871, A NOVEL HUMAN G PROTEIN COUPLED RECEPTOR AND USES THEREOF**

(57) Abstract: The invention provides isolated nucleic acids molecules, designated 52871 nucleic acid molecules, which encode novel G-protein Coupled Receptor molecules. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing 52871 nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which a 52871 gene has been introduced or disrupted. The invention still further provides isolated 52871 proteins, fusion proteins, antigenic peptides and anti-52871 antibodies. Diagnostic methods utilizing compositions of the invention are also provided.

WO 01/96400 A2

## 52871, A NOVEL HUMAN G PROTEIN COUPLED RECEPTOR AND USES THEREOF

### Related Applications

- 5 This application claims priority to U.S. Provisional Patent Application No. 60/212,331, filed on June 16, 2000, and to U.S. Provisional Patent Application No. 60/269,758, filed on February 16, 2001, the contents of which are incorporated herein in their entirety by reference.

### 10 Background of the Invention

- Molecular cloning studies have shown that G protein-coupled receptors ("GPCRs") form one of the largest protein superfamilies found in nature, and it is estimated that greater than 1000 different such receptors exist in mammals. Upon binding of extracellular ligands, GPCRs interact with a specific subset of heterotrimeric G-proteins that can then, in their activated forms, inhibit or activate various effector enzymes and/or ion channels. The ligands for many of these receptors are known although there exists an ever-increasing number of GPCRs which have been identified in the sequencing of the human genome for which no ligands have yet been identified. This latter subfamily of GPCRs is called the orphan family of GPCRs. In addition to both GPCRs with known ligands, as well as orphan GPCRs, there exist a family of GPCR-like molecules which share significant homology as well as many of the structural properties of the GPCR superfamily. For example, a family of GPCR-like proteins which arises from three alternatively-spliced forms of a gene occurring between the CD4 and triosephosphate isomerase genes at human chromosome 12p13, has been recently identified (Ansari-Lari *et al.* (1996) *Genome Res.* 6:314-326). Comparative sequence analysis of the syntenic region in mouse chromosome 6 has further revealed a murine homologue of one of these GPCR splice variants (Ansari-Lari *et al.* (1998) *Genome Res.* (1):29-40.

- The fundamental knowledge that GPCRs play a role in regulating activities in virtually every cell in the human body has fostered an extensive search for modulators of such receptors for use as human therapeutics. In fact, the superfamily of GPCRs has proven to be among the most successful drug targets. Consequently, it has been recognized that the newly isolated orphan GPCRs, as well as the GPCR-like proteins, have great potential for drug discovery. With the identification of each new GPCR, orphan GPCR, and GPCR-like proteins, there exists a need for identifying the surrogate ligands for such molecules as well as for modulators of such molecules for use in regulating a variety of cellular responses.

### Summary of the Invention

The present invention is based, at least in part, on the discovery of novel G Protein Coupled Receptor family members, referred to herein as "52871" nucleic acid and protein molecules. The 52871 nucleic acid and protein molecules of the present invention (as well as modulators of said 52871 nucleic acid and protein molecules) are useful as agents in regulating a variety of cellular processes, *e.g.*, cellular proliferation, growth, differentiation, nociception, and signaling (*e.g.*, pain signaling). The 52871 nucleic acid and protein molecules (and modulators thereof) are also useful in regulating physiologic processes, for example, pain and/or pain disorders. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding 52871 proteins or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of 52871-encoding nucleic acids.

In one embodiment, the invention features an isolated nucleic acid molecule that includes the nucleotide sequence set forth in SEQ ID NO:1 or SEQ ID NO:3. In another embodiment, the invention features an isolated nucleic acid molecule that encodes a polypeptide including the amino acid sequence set forth in SEQ ID NO:2. In another embodiment, the invention features an isolated nucleic acid molecule that includes the nucleotide sequence contained in the plasmid deposited with ATCC® as Accession Number \_\_\_\_\_.

In still other embodiments, the invention features isolated nucleic acid molecules including nucleotide sequences that are substantially identical (*e.g.*, 60% identical) to the nucleotide sequence set forth as SEQ ID NO:1 or SEQ ID NO:3. The invention further features isolated nucleic acid molecules including at least 30 contiguous nucleotides of the nucleotide sequence set forth as SEQ ID NO:1 or SEQ ID NO:3. In another embodiment, the invention features isolated nucleic acid molecules which encode a polypeptide including an amino acid sequence that is substantially identical (*e.g.*, 60% identical) to the amino acid sequence set forth as SEQ ID NO:2. Also features are nucleic acid molecules which encode allelic variants of the polypeptide having the amino acid sequence set forth as SEQ ID NO:2. In addition to isolated nucleic acid molecules encoding full-length polypeptides, the present invention also features nucleic acid molecules which encode fragments, for example biologically active or antigenic fragments, of the full-length polypeptides of the present invention (*e.g.*, fragments including at least 10 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:2). In still other embodiments, the invention features nucleic acid molecules that are complementary to, are antisense to, or hybridize under stringent conditions to the isolated nucleic acid molecules described herein.

In a related aspect, the invention provides vectors including the isolated nucleic acid molecules described herein (*e.g.*, 52871-encoding nucleic acid molecules). Such vectors can optionally include nucleotide sequences encoding heterologous polypeptides. Also featured are host cells including such vectors (*e.g.*, host cells  
5 including vectors suitable for producing 52871 nucleic acid molecules and polypeptides).

In another aspect, the invention features isolated 52871 polypeptides and/or biologically active or antigenic fragments thereof. Exemplary embodiments feature a polypeptide including the amino acid sequence set forth as SEQ ID NO:2, a polypeptide  
10 including an amino acid sequence at least 60% identical to the amino acid sequence set forth as SEQ ID NO:2, a polypeptide encoded by a nucleic acid molecule including a nucleotide sequence at least 60% identical to the nucleotide sequence set forth as SEQ ID NO:1 or SEQ ID NO:3. Also featured are fragments of the full-length polypeptides described herein (*e.g.*, fragments including at least 10 contiguous amino acid residues of  
15 the sequence set forth as SEQ ID NO:2) as well as fragments of allelic variants of the polypeptide having the amino acid sequence set forth as SEQ ID NO:2.

The 52871 polypeptides and/or biologically active or antigenic fragments thereof, are useful, for example, as reagents or targets in assays applicable to treatment and/or diagnosis of 52871 mediated or related disorders (*e.g.*, pain disorders). In one  
20 embodiment, a 52871 polypeptide or fragment thereof has a 52871 activity. In another embodiment, a 52871 polypeptide or fragments thereof has a transmembrane domain, and/or a "7 transmembrane receptor profile" and optionally, has a 52871 activity. In a related aspect, the invention features antibodies (*e.g.*, antibodies which specifically bind to any one of the polypeptides, as described herein) as well as fusion polypeptides  
25 including all or a fragment of a polypeptide described herein.

The present invention further features methods for detecting 52871 polypeptides and/or 52871 nucleic acid molecules, such methods featuring, for example, a probe, primer or antibody described herein. Also featured are kits for the detection of 52871 polypeptides and/or 52871 nucleic acid molecules. In a related aspect, the invention  
30 features methods for identifying compounds which bind to and/or modulate the activity of a 52871 polypeptide or 52871 nucleic acid molecule described herein. Also featured are methods for modulating a 52871 activity.

The present invention also provides diagnostic assays for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant  
35 modification or mutation of a gene encoding a 52871 protein; (ii) mis-regulation of the gene; and (iii) aberrant post-translational modification of a 52871 protein, wherein a wild-type form of the gene encodes a protein with a 52871 activity.



The invention further provides methods for identifying 52871 modulators. In one embodiment, the invention provides a method for identifying a compound which binds to a 52871 polypeptide by contacting the polypeptide, or a cell expressing the polypeptide with a test compound, and determining whether the polypeptide binds to the test compound. In another embodiment, the invention provides a method for identifying a compound which modulates the activity of a 52871 polypeptide comprising contacting a 52871 polypeptide or a cell which expresses a 52871 polypeptide with a test compound and determining the effect of the test compound on the activity of the polypeptide. In another embodiment, the 52871 activity is modulation of nociception. In yet another embodiment, the 52871 activity is modulation of pain signaling.

Accordingly, in a further aspect, the invention provides a method for identifying a compound which modulates pain comprising contacting a 52871 polypeptide or a cell which expresses a 52871 polypeptide with a test compound with a test compound and identifying the compound as a modulator of pain by determining the effect of the test compound on the activity of the polypeptide. In yet another aspect, the invention provides a method for identifying a compound capable of modulating nociception comprising contacting a 52871 polypeptide or a cell which expresses a 52871 polypeptide with a test compound and identifying the compound as a modulator of nociception by determining the effect of the test compound on the activity of the polypeptide.

The present invention further features a method for treating a subject having pain or a pain disorder comprising administering to the subject a 52871 modulator. In one embodiment, the 52871 modulator is a small molecule. In another embodiment, the 52871 modulator is administered in a pharmaceutically acceptable formulation. In yet another embodiment the 52871 modulator is administered using a gene therapy vector.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

#### **Brief Description of the Drawings**

*Figures 1A-1B* depict the cDNA sequence and predicted amino acid sequence of human 52871 (clone Fbh52871). The nucleotide sequence corresponds to nucleic acids 1 to 1731 of SEQ ID NO:1. The amino acid sequence corresponds to amino acids 1 to 348 of SEQ ID NO:2. The coding region without the 3' untranslated region of the human 52871 gene is shown in SEQ ID NO:3.

*Figure 2* depicts a structural, hydrophobicity, and antigenicity analysis of the human Fbh52871 protein.

*Figure 3* depicts the results of a search which was performed against the HMM database and which resulted in the identification of a 7 transmembrane receptor profile (indicated as "7tm\_1").

### **Detailed Description of the Invention**

The present invention is based, at least in part, on the discovery of novel G Protein Coupled Receptor (GPCR) molecules, referred to herein as "52871" nucleic acid and protein molecules, which are novel members of a family of receptors which possess the ability to associate with G protein molecules in order to function in their biological capacity (*e.g.*, to modulate target enzymes or ion channels). These novel molecules are capable of participating in signaling pathways (*e.g.*, as a hormone receptor, as a neurotransmitter receptor, as an modulator of intracellular signaling) and, thus, play a role in or function in a variety of cellular processes, *e.g.*, cellular signaling. GPCRs act as the receptors for various different families of neuropeptides (Pheng and Regoli (2000) *Life Sci* 67:847; Rozengurt, E. (1998) *J. Cell Physiology* 177:507; Larhammar, *et al.* (1993) *Drug Des. Discovery* 9:179; Elshourbagy, *et al.* (2000) *J Biol Chem* 275:25965). Neuropeptides are known to be involved in nociception (*e.g.*, chemical, mechanical, or thermal nociception), and thereby function to modulate pain elicitation (Bannon *et al.* (2000) *Brain Res* 868:79). 52871 mRNA is predominantly expressed in the brain, spinal cord, dorsal root ganglia (DRG), and skin, as compared to other tissues. Based at least in part on the specific expression pattern of 52871, in addition to the fact that GPCRs are known to be involved in the modulation of nociceptive neurons and pain signaling (through interaction with neuropeptides), the novel 52871 molecules of the present invention can be used as targets for developing novel diagnostic targets and therapeutic agents to control pain and pain disorders.

As used herein, the term "G protein coupled receptor" (referred to herein interchangeably as "GPCR") includes a protein, peptide, or enzyme which is able to interact with one or more G protein molecules (*e.g.*, neuropeptides) and or one or more signaling molecules, in order to carry out its function(s), *e.g.*, recognition of signaling molecules, modulation of intracellular signaling, modulation of nociception, and/or modulation of pain. GPCR molecules are involved in the transduction of signals that are transmitted to cells from without by important signaling molecules, including peptides, hormones, growth factors, and neurotransmitters. GPCR proteins act as cell-surface receptors for these molecules, with an extracellular domain which interacts with the signaling molecule, and an intracellular domain which can further activate signaling events. The activity of the intracellular domain is typically sensitive to the binding state

of the extracellular domain (*e.g.*, ligand bound, ligand unbound). Normally the intracellular domain can interact with guanine nucleotide-binding (G) proteins, thus activating it. Thereupon, the activated G protein can modulate the activity of one or more enzymes. This enzyme can pass the signal on directly by catalyzing the  
5 production of a second messenger (*e.g.*, cyclic AMP; cAMP), or it can catalyze the production of a soluble mediator (*e.g.*, inositol triphosphate; IP<sub>3</sub>) which can in turn release other second messengers (*e.g.*, Ca<sup>2+</sup> from the endoplasmic reticulum). Examples of GPCR molecules include prokaryotic, plant, and mammalian GPCR molecules. As transmembrane receptor protein, the 52871 molecules of the present invention provide  
10 novel diagnostic targets and therapeutic agents to control GPCR-associated disorders.

Preferably such GPCR proteins comprise a family of GPCR molecules. The term "family" when referring to the protein and nucleic acid molecules of the invention is intended to mean two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence  
15 homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin, as well as other, distinct proteins of human origin or alternatively, can contain homologues of non-human origin, *e.g.*, mouse or monkey proteins. Members of a family also have common functional characteristics.

20 For example, the family of G protein-coupled receptors (GPCRs), to which the 52871 proteins of the present invention bear significant homology, comprise an N-terminal extracellular domain, seven transmembrane domains (also referred to as membrane-spanning domains), three extracellular domains (also referred to as extracellular loops), three cytoplasmic domains (also referred to as cytoplasmic loops),  
25 and a C-terminal cytoplasmic domain (also referred to as a cytoplasmic tail). Members of the GPCR family also share certain conserved amino acid residues, some of which have been determined to be critical to receptor function and/or G protein signaling. For example, GPCRs contain the following features: a conserved asparagine residue in the first transmembrane domain; a cysteine residue in the first extracellular loop which is  
30 believed to form a disulfide bond with a conserved cysteine residue in the second extracellular loop; a conserved leucine and aspartate residue in the second transmembrane domain; an aspartate-arginine-tyrosine motif (DRY motif) at the interface of the third transmembrane domain and the second cytoplasmic loop of which the arginine residue is almost invariant (members of the rhodopsin subfamily of GPCRs  
35 comprise a histidine-arginine-methionine motif (HRM motif) as compared to a DRY motif); a conserved tryptophan and proline residue in the fourth transmembrane domain; a conserved phenylalanine residue which is commonly found as part of the motif FXXCXXP; and a conserved leucine residue in the seventh transmembrane domain

which is commonly found as part of the motif DPXXY or NPXXY. Table I depicts an alignment of the seven transmembrane domains (TM1 – TM7) of 5 known GPCRs. The conserved residues described herein are indicated by asterices.

5 TABLE I ALIGNMENT OF:

	Name	SEQ ID	species	GenBank Acc No.				
	thrombin	(4)	human	P25116				
	rhodopsin	(5)	human	P08100				
10	m1ACh	(6)	rat	P08482				
	IL-8A	(7)	human	P25024				
	octopamine	(8)	Drosophila melanogaster	P22270				
15	TM1							
			*					
	4.	102	TLFVPSVYTG	VFVVS	LPNIMAIVVFILKMK	132		
	5.	37	*FSMLAAYM	FLLIVL	GFPINFLTL	VTVQHKK	67	
20	6.	25	VAFIGITT	GLLSLAT	VTGNLLVL	ISFKVNTE	55	
	7.	39	KYVVIIAY	ALVFLS	LLGNSLV	MLVILYSRV	69	
	8.	109	ALLTALVLS	SVIIVL	TIIGNILV	ILSVFTYKP	139	
25	TM2							
			*	*				
	4.	138	VVYMLHLA	TADVLF	VS	LPFKISYYFSG	165	
	5.	73	NYILLNL	AVADLF	FMVLGGFT	STLYTSLH	100	
	6.	61	NYFLLSL	ACADLI	IGTF	SMNLYT	TYLLM	88
30	7.	75	DVYLLNL	ALADLL	FALTLP	IWAASKVNG	102	
	8.	145	NFFIVSL	AVADLT	VALLVL	PFNVAYSIL	172	
35	TM3							
				*				
	4.	176	RFVTA	AFYCNMY	ASILLMT	VISIDR	200	
	5.	111	NLEGFF	FATLGGE	IALWSL	VVLAIER	135	
	6.	99	DLWLAL	DYVASN	ASVMNLL	LISFDR	123	
40	7.	111	KVVSLL	KEVNFY	SGILL	LACISVDR	135	
	8.	183	KLWLTC	DVLCCT	SSILNL	CAIALDR	207	
45	TM4							
			*	*				
	4.	215	TLGRAS	FTCLAI	WALAIAG	VVPLVLKE	241	
	5.	149	GENHAI	MGVAFT	WVMALAC	AAPPLAGW	175	
	6.	138	TPRRAA	LMIGLAW	LVSFVL	WAPAILFW	164	
	7.	149	KRHLVK	FVCLGC	WGLSMNL	SLPFFLFR	175	
50	8.	222	TVGRVLL	LISGVW	LLSLLIS	SPPLIGW	248	
	TM5							
			*	*	*	*		
55	4.	268	AYYFSA	FAFSAV	FFVPLI	IISTVCYVSIIRC	296	

5.	201	ESFVIYMFVVHFTIPMIIFFCYGQLVFT	229
6.	186	PIITFGTAMAAFYLPVTVMCTLYWRIYRE	214
7.	200	MVLRILPHTFGFIVPLFVMLFCYGFTRLRT	228
8.	267	RGYVIYSSLGSFFIPLAIMTIVYIEIFVA	295
5			
TM6			
* * *			
4.	313	FLSAAVFCIFIICFGPTNVLLIAHYSFL	340
10	5.	252 RMVIIMVIAFLICWVPYASVAFYIFTHQ	279
	6.	365 RTLSAILLAFILTWTPYNIMVLVSTFCK	397
	7.	242 RVIFAVVLIFLLCWLPLYNLVLLADTLMR	269
	8.	529 RTLGIIMGVFCVLCWLPFFLMYVILPFCQ	556
15			
TM7			
** *			
4.	347	EAAVFAYLLCVCVSSISSCIDPLIYYYASSECQ	379
5.	282	NFGPIFMTIPAFFAKSAAIYNPVIYIMMNKQFR	314
6.	394	CVPETLWELGYWLCYVNSTVNPMCYALCNKAFR	426
20	7.	281 NNIGRALDATEILGFLHSCLNPIIYAFIQNFR	313
	8.	559 CPTNKFKNFITWLGYNISGLNPVIYTIFNLDYR	591

The amino acid sequences of thrombin (GenBank Accession No. 25116), rhodopsin (GenBank Accession No. 08100), m1ACh (GenBank Accession No. 08482), IL-8A (GenBank Accession No. 25024), octopamine (GenBank Accession No. 22270), can be found as SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, and SEQ ID NO:8, respectively.

Accordingly, GPCR-like proteins such as the 52871 proteins of the present invention contain a significant number of structural characteristics of the GPCR family. For instance, the 52871 proteins of the present invention contain conserved cysteines found in the first 2 extracellular loops (prior to the third and fifth transmembrane domains) of most GPCRs (cys 121 and cys 197 of SEQ ID NO:2). A highly conserved asparagine residue in the first transmembrane domain is present (asn 67 in SEQ ID NO:2). Transmembrane domain two of the 52871 proteins contains a highly conserved leucine (leu 90 of SEQ ID NO:2). The two cysteine residues are believed to form a disulfide bond that stabilizes the functional protein structure. A highly conserved tryptophan and proline in the fourth transmembrane domain of the 52871 proteins is present (trp171 and pro 180 of SEQ ID NO:2). The third cytoplasmic loop contains 40 amino acid residues and is thus the longest cytoplasmic loop of the three, characteristic of G protein coupled receptors. Moreover, a highly conserved proline in the sixth transmembrane domain is present (pro 289 of SEQ ID NO:2). The proline residues in the fourth, fifth, sixth, and seventh transmembrane domains are thought to introduce kinks in the alpha-helices and may be important in the formation of the ligand binding pocket. Moreover, an almost invariant proline is present in the seventh transmembrane domain of 52871 (pro327 of SEQ ID NO:2).

In one embodiment, the 52871 proteins of the present invention are proteins having an amino acid sequence of about 200-475, preferably about 250-425, more preferably about 275-400, more preferably about 300-375, or about 330-350 amino acids in length. In another embodiment, the 52871 proteins of the present invention contain at least one transmembrane domain. As used herein, the term "transmembrane domain" includes an amino acid sequence having at least about 10, preferably about 13, preferably about 16, more preferably about 19, 21, 23, 25, 30, 35 or 40 amino acid residues, of which at least about 50-60%, 60-70%, preferably about 70-80% more preferably about 80-90%, or about 90-95% of the amino acid residues contain non-polar side chains, for example, alanine, valine, leucine, isoleucine, proline, phenylalanine, tryptophan, and methionine. A transmembrane domain is lipophilic in nature. Transmembrane domains are described in, for example, Zagotta W.N. *et al.*, (1996) *Annual Rev. Neurosci.* 19:235-63, the contents of which are incorporated herein by reference. For example, a transmembrane domain can be found at about amino acids 53-75 of SEQ ID NO:2. In a preferred embodiment, a 52871 protein of the present invention has more than one transmembrane domain, preferably 2, 3, 4, 5, 6, or 7 transmembrane domains. For example, transmembrane domains can be found at about amino acids 53-75, 90-108, 126-144, 165-186, 210-234, 275-293, and 309-333 of SEQ ID NO:2. In a particularly preferred embodiment, a 52871 protein of the present invention has 7 transmembrane domains.

In another embodiment, a 52871 family member is identified based on the presence of at least one cytoplasmic loop, also referred to herein as a cytoplasmic domain. In another embodiment, a 52871 family member is identified based on the presence of at least one extracellular loop. As defined herein, the term "loop" includes an amino acid sequence having a length of at least about 4, preferably about 5-10, preferably about 10-20, and more preferably about 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, or 100-150 amino acid residues, and has an amino acid sequence that connects two transmembrane domains within a protein or polypeptide. Accordingly, the N-terminal amino acid of a loop is adjacent to a C-terminal amino acid of a transmembrane domain in a naturally-occurring GPCR or GPCR-like molecule, and the C-terminal amino acid of a loop is adjacent to an N-terminal amino acid of a transmembrane domain in a naturally-occurring GPCR or GPCR-like molecule.

As used herein, a "cytoplasmic loop" includes an amino acid sequence located within a cell or within the cytoplasm of a cell. For example, a cytoplasmic loop is found at about amino acids 76-89, 145-164, and 235-274 of SEQ ID NO:2. Also as used herein, an "extracellular loop" includes an amino acid sequence located outside of a cell, or extracellularly. For example, an extracellular loop can be found at about amino acid residues 109-125, 187-209, and 294-308 of SEQ ID NO:2.

In another embodiment of the invention, a 52871 family member is identified based on the presence of a "C-terminal cytoplasmic domain", also referred to herein as a C-terminal cytoplasmic tail, in the sequence of the protein. As used herein, a "C-terminal cytoplasmic domain" includes an amino acid sequence having a length of at least 5 amino acid residues and is located within a cell or within the cytoplasm of a cell. Accordingly, the N-terminal amino acid residue of a "C-terminal cytoplasmic domain" is adjacent to a C-terminal amino acid residue of a transmembrane domain in a naturally-occurring GPCR or GPCR-like protein. For example, a C-terminal cytoplasmic domain is found at about amino acid residues 334-348 of SEQ ID NO:5.

10 In another embodiment, a 52871 family member is identified based on the presence of an "N-terminal extracellular domain", also referred to herein as an N-terminal extracellular loop in the amino acid sequence of the protein. As used herein, an "N-terminal extracellular domain" includes an amino acid sequence having about 1-500, preferably about 1-400, more preferably about 1-300, more preferably about 1-200, even  
15 more preferably about 1-100, and even more preferably about 1-60 amino acid residues in length and is located outside of a cell or extracellularly. The C-terminal amino acid residue of a "N-terminal extracellular domain" is adjacent to an N-terminal amino acid residue of a transmembrane domain in a naturally-occurring GPCR or GPCR-like protein. For example, an N-terminal cytoplasmic domain is found at about amino acid  
20 residues 1-52 of SEQ ID NO:2.

Accordingly in one embodiment of the invention, a 52871 family member includes at least one transmembrane domain and/or at least one cytoplasmic loop, and/or at least one extracellular loop. In another embodiment, the 52871 family member further includes an N-terminal extracellular domain and/or a C-terminal cytoplasmic  
25 domain. In another embodiment, the 52871 family member can include up to six transmembrane domains, three cytoplasmic loops, and two extracellular loops, or can include up to six transmembrane domains, three extracellular loops, and two cytoplasmic loops. The former embodiment can further include an N-terminal extracellular domain. The latter embodiment can further include a C-terminal  
30 cytoplasmic domain. In another embodiment, the 52871 family member can include seven transmembrane domains, three cytoplasmic loops, and three extracellular loops and can further include an N-terminal extracellular domain or a C-terminal cytoplasmic domain.

In another embodiment, a 52871 family member is identified based on the  
35 presence of at least one "7 transmembrane receptor profile", also referred to as a "7-TMR profile", in the protein or corresponding nucleic acid molecule. As used herein, the term "7-TMR profile" includes an amino acid sequence having at least about 100-400, preferably about 150-350, more preferably about 200-300 amino acid residues, or

at least about 250-260 amino acids in length and having a bit score for the alignment of the sequence to the "7tm\_1" family Hidden Markov Model (HMM) of at least 100, preferably 100-110, more preferably 110-120, more preferably 120-130, 130-140, 140-150, 150-160 or greater. The 7tm\_1 family HMM has been assigned the PFAM

5 Accession PF00001 (<http://www.sanger.ac.uk/Software/Pfam/>).

To identify the presence of a 7-TMR profile in a 52871 family member, the amino acid sequence of the protein family member is searched against a database of HMMs (*e.g.*, the Pfam database, release 2.1) using the default parameters ([http://www.sanger.ac.uk/Software/Pfam/HMM\\_search](http://www.sanger.ac.uk/Software/Pfam/HMM_search)). For example, the search can be  
 10 performed using the hmmsf program (family specific) using the default parameters (*e.g.*, a threshold score of 15) for determining a hit. hmmsf is available as part of the HMMER package of search programs (HMMER 2.1.1, Dec. 1998) which is freely distributed by the Washington University School of Medicine. Alternatively, the threshold score for determining a hit can be lowered (*e.g.*, to 8 bits). For example, a  
 15 search using the amino acid sequence of SEQ ID NO:2 was performed against the HMM database resulting in the identification of a 7tm\_1 receptor profile in the amino acid sequence of human 52871 (SEQ ID NO:2) at about residues 66-330 having a score of 164.0. The results of the search are set forth in Figure 3.

A description of the Pfam database can be found in Sonhammer *et al.* (1997)  
 20 *Proteins* 28:405-420 and a detailed description of HMMs can be found, for example, in Gribskov *et al.* (1990) *Meth. Enzymol.* 183:146-159; Gribskov *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.* (1994) *J. Mol. Biol.* 235:1501-1531; and Stultz *et al.* (1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference.

25 In a preferred embodiment, a 52871 molecule, as described herein is characterized by the presence of a "G-protein coupled receptor signature." As used herein, the term "G-protein coupled receptor signature" includes a motif having the consensus sequence [GSTALIVMFYWC]-[GSTANCPDE]-{EDPKRH}-X(2)-[LIVMNQGA]-X(2)-[LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-  
 30 [FYWCSH]-X(2)-[LIVM] and is described under Prosite entry number PDOC00210 (<http://www.expasy.ch/prosite>). A G-protein coupled receptor signature can be found, for example, within the 7-TMR profile of the 52871 protein of SEQ ID NO:2 at about residues 134-150. The consensus sequences described herein are described according to standard Prosite Signature designation (*e.g.*, all amino acids are indicated according to  
 35 their universal single letter designation; X designates any amino acid; X(n) designates any n amino acids, *e.g.*, X(2) designates any 2 amino acids; [LIVM] indicates any one of the amino acids appearing within the brackets, *e.g.*, any one of L, I, V, or M, in the alternative, any one of Leu, Ile, Val, or Met.); and {LIVM} indicates any amino acid



except the amino acids appearing within the brackets, *e.g.*, not L, not I, not V, and not M.

Isolated proteins of the present invention, for example 52871 proteins, preferably have an amino acid sequence sufficiently identical to the amino acid sequence of SEQ ID NO:2, or are encoded by a nucleotide sequence sufficiently identical to SEQ ID NO:1 or 3. As used herein, the term "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (*e.g.*, an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains or motifs and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homology or identity across the amino acid sequences of the domains and contain at least one and preferably two structural domains or motifs, are defined herein as sufficiently identical. Furthermore, amino acid or nucleotide sequences which share at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homology or identity and share a common functional activity are defined herein as sufficiently identical.

In a preferred embodiment, a 52871 protein includes at least one or more of the following domains: a transmembrane domain, and/or a 7-TMR profile, and has an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homologous or identical to the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_. In yet another preferred embodiment, a 52871 protein includes at least one or more of the following domains: a transmembrane domain, a 7-TMR profile, and is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or 3. In another preferred embodiment, a 52871 protein includes at least one or more of the following domains: a transmembrane domain, a 7-TMR profile, and has a 52871 activity.

As used interchangeably herein, a "52871 activity", "biological activity of 52871" or "functional activity of 52871", refers to an activity exerted by a 52871 protein, polypeptide or nucleic acid molecule on a 52871 responsive cell as determined *in vivo*, or *in vitro*, according to standard techniques. In one embodiment, a 52871 activity is a direct activity, such as an association with a 52871-target molecule. As used herein, a "target molecule" or "binding partner" or "ligand" or "substrate" is a molecule with which a 52871 protein binds

or interacts in nature, such that 52871-mediated function is achieved. A 52871 target molecule can be a non-52871 molecule or a 52871 protein or polypeptide of the present invention. In an exemplary embodiment, a 52871 target molecule is a 52871 ligand such as hormone, a neurotransmitter, a growth factor, an opioid, a pheromone, a peptide (*e.g.*, a cytokine, a chemokine, a neuropeptide), a biogenic amine, an eicosanoid, a lipid (*e.g.*, a leukotriene, a cannabinoids), an excitatory amino acid (*e.g.*, GABA, glutamate), an ion (*e.g.*, calcium), or a retinoid). Examples of 52871 substrates also include molecules that are essential for 52871 intracellular function, *e.g.*, G protein, adenylyl cyclase, enzymes involved in the inositol triphosphate signaling pathways. Alternatively, a 52871 activity is an indirect activity, such as a cellular signaling activity (*e.g.*, ligand recognition, modulation of intracellular signaling mechanisms, modulation of nociception, modulation of pain) mediated by interaction of the 52871 protein with a 52871 ligand.

In a preferred embodiment, a 52871 activity is at least one or more of the following activities: (i) interaction of a 52871 protein with soluble 52871 ligand; (ii) interaction of a 52871 protein with a membrane-bound non-52871 protein; (iii) interaction of a 52871 protein with an intracellular protein (*e.g.*, an intracellular enzyme or signal transduction molecule); (iv) indirect interaction of a 52871 protein with an intracellular protein (*e.g.*, a downstream signal transduction molecule); (v) modulation of intra- or intercellular signaling or feedback mechanisms; (vi) modulation of the intracellular levels and/or homeostatic balance of signaling molecule pools (*e.g.*,  $\text{Ca}^{2+}$ , diacylglycerol,  $\text{IP}_3$ , cAMP); (vii) regulation of cellular proliferation; (viii) regulation of cellular differentiation; (ix) regulation of development; (x) regulation of gene expression in a cell which expresses a 52871 protein; (xi) modulation of nociception; (xii) modulation of pain; and (xiii) regulation of cell death.

Accordingly, another embodiment of the invention features isolated 52871 proteins and polypeptides having a 52871 activity. Other preferred proteins are 52871 proteins having one or more of the following domains: a transmembrane domain, a 7-TMR profile, and, preferably, a 52871 activity.

Additional preferred proteins have at least one transmembrane domain, one 7-TMR profile, and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or 3.

The nucleotide sequence of the isolated human 52871 cDNA and the predicted amino acid sequence of the human 52871 polypeptide are shown in Figure 1 and in SEQ ID NOs:1 and 2, respectively. A plasmid containing the nucleotide sequence encoding human 52871 was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, USA, on \_\_\_\_\_ and

assigned Accession Number \_\_\_\_\_. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

The human 52871 gene, which is approximately 1731 nucleotides in length and encodes a protein which is approximately 348 amino acid residues in length.

Various aspects of the invention are described in further detail in the following subsections:

#### I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode 52871 proteins or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify 52871-encoding nucleic acid molecules (*e.g.*, 52871 mRNA) and fragments for use as PCR primers for the amplification or mutation of 52871 nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

The term "isolated nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated 52871 nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_, or a portion thereof, can be isolated using standard molecular biology techniques and the

sequence information provided herein. Using all or portion of the nucleic acid sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_ as a hybridization probe, 52871 nucleic acid molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_ can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_.

A nucleic acid of the invention can be amplified using cDNA, mRNA or, alternatively, genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to 52871 nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In one embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:1 or 3. This cDNA may comprise sequences encoding the human 52871 protein (i.e., "the coding region", from nucleotides 201-1247), as well as 5' untranslated sequences (nucleotides 1-200) and 3' untranslated sequences (nucleotides 1248-1731) of SEQ ID NO:1. Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:1 (e.g., nucleotides 201-1247, corresponding to SEQ ID NO:3). Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention comprises SEQ ID NO:3 and nucleotides 1-200 of SEQ ID NO:1. In yet another embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:3 and nucleotides 1248-1731 of SEQ ID NO:1. In yet another embodiment, the nucleic acid molecule consists of the nucleotide sequence set forth as SEQ ID NO:1 or SEQ ID NO:3.

In still another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, is one

which is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_ such that it can hybridize to a complement of the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, respectively, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to the nucleotide sequence shown in SEQ ID NO:1 or 3 (*e.g.*, to the entire length of the nucleotide sequence), or to the nucleotide sequence (*e.g.*, the entire length of the nucleotide sequence) of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or to a portion or complement of any of these nucleotide sequences. In one embodiment, a nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least (or no greater than) 50-100, 100-250, 250-500, 500-750, 750-1000, 1000-1250, 1250-1500, 1500-1700 or more nucleotides in length and hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, for example, a fragment which can be used as a probe or primer or a fragment encoding a portion of a 52871 protein, *e.g.*, a biologically active portion of a 52871 protein. The nucleotide sequence determined from the cloning of the 52871 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other 52871 family members, as well as 52871 homologues from other species. The probe/primer typically comprises substantially purified oligonucleotide. The probe/primer (*e.g.*, oligonucleotide) typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to a complement of at least about 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_ of an anti-sense sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_ or of a naturally occurring allelic variant or mutant of SEQ ID NO:1 or 3, or the nucleotide

sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_.

Exemplary probes or primers are at least (or no greater than) 12 or 15, 20 or 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75 or more nucleotides in length and/or comprise  
5 consecutive nucleotides of an isolated nucleic acid molecule described herein. Also included within the scope of the present invention are probes or primers comprising contiguous or consecutive nucleotides of an isolated nucleic acid molecule described herein, but for the difference of 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 bases within the probe or primer sequence. Probes based on the 52871 nucleotide sequences can be used to detect  
10 (*e.g.*, specifically detect) transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. In another embodiment a set of primers is provided, *e.g.*, primers suitable for use in a PCR, which can be used to amplify a  
15 selected region of a 52871 sequence, *e.g.*, a domain, region, site or other sequence described herein. The primers should be at least 5, 10, or 50 base pairs in length and less than 100, or less than 200, base pairs in length. The primers should be identical, or differs by no greater than 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 bases when compared to a sequence disclosed herein or to the sequence of a naturally occurring variant. Such  
20 probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a 52871 protein, such as by measuring a level of a 52871-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting 52871 mRNA levels or determining whether a genomic 52871 gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of a 52871  
25 protein" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_ which encodes a polypeptide having a 52871 biological activity (the biological activities of the 52871 proteins are described herein), expressing the encoded portion of the 52871 protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the 52871 protein. In an  
30 exemplary embodiment, the nucleic acid molecule is at least 50-100, 100-250, 250-500, 500-700, 750-1000, 1000-1250, 1250-1500, 1500-1700 or more nucleotides in length and encodes a protein having a 52871 activity (as described herein).

The invention further encompasses nucleic acid molecules that differ from the  
35 nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_. Such differences can be due to degeneracy of the genetic code, thus resulting in a nucleic acid which encodes the same 52871 proteins as those encoded by the nucleotide sequence

shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence which differs by at least 1, but no greater than 5,  
5 10, 20, 50 or 100 amino acid residues from the amino acid sequence shown in SEQ ID NO:2, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession Number \_\_\_\_\_. In yet another embodiment, the nucleic acid molecule encodes the amino acid sequence of human 52871. If an alignment is needed for this comparison, the sequences should be aligned for maximum homology.

10 Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologues (different locus), and orthologues (different organism) or can be non naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can  
15 occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product).

Allelic variants result, for example, from DNA sequence polymorphisms within a population (*e.g.*, the human population) that lead to changes in the amino acid  
20 sequences of the 52871 proteins. Such genetic polymorphism in the 52871 genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding a 52871 protein, preferably a mammalian 52871 protein, and can further include non-coding regulatory sequences, and introns.

25 Accordingly, in one embodiment, the invention features isolated nucleic acid molecules which encode a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, wherein the nucleic acid molecule hybridizes to a complement of a nucleic acid molecule comprising  
30 SEQ ID NO:1 or SEQ ID NO:3, for example, under stringent hybridization conditions.

Allelic variants of human 52871 include both functional and non-functional 52871 proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the human 52871 protein that maintain the ability to process a 52871 substrate (*e.g.*, carboxylation, decarboxylation). Functional allelic variants will typically  
35 contain only conservative substitution of one or more amino acids of SEQ ID NO:2, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein.

Non-functional allelic variants are naturally occurring amino acid sequence variants of the human 52871 protein that do not have the ability to bind or process a 52871 substrate (*e.g.*, signal molecule recognition, signal transduction), and/or carry out any of the 52871 activities described herein. Non-functional allelic variants will  
5 typically contain a non-conservative substitution, a deletion, or insertion or premature truncation of the amino acid sequence of SEQ ID NO:2, or a substitution, insertion or deletion in critical residues or critical regions of the protein.

The present invention further provides non-human orthologues (*e.g.*, non-human orthologues of the human 52871 protein). Orthologues of the human 52871 protein are  
10 proteins that are isolated from non-human organisms and possess the same 52871 substrate binding and/or modulation of membrane excitability activities of the human 52871 protein. Orthologues of the human 52871 protein can readily be identified as comprising an amino acid sequence that is substantially identical to SEQ ID NO:2.

Moreover, nucleic acid molecules encoding other 52871 family members and,  
15 thus, which have a nucleotide sequence which differs from the 52871 sequences of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_ are intended to be within the scope of the invention. For example, another 52871 cDNA can be identified based on the nucleotide sequence of human 52871. Moreover, nucleic acid molecules encoding 52871 proteins  
20 from different species, and which, thus, have a nucleotide sequence which differs from the 52871 sequences of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_ are intended to be within the scope of the invention. For example, a mouse or monkey 52871 cDNA can be identified based on the nucleotide sequence of a human 52871.

25 Nucleic acid molecules corresponding to natural allelic variants and homologues of the 52871 cDNAs of the invention can be isolated based on their homology to the 52871 nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Nucleic acid molecules corresponding to natural  
30 allelic variants and homologues of the 52871 cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the 52871 gene.

Orthologues, homologues and allelic variants can be identified using methods known in the art (*e.g.*, by hybridization to an isolated nucleic acid molecule of the present invention, for example, under stringent hybridization conditions). In one  
35 embodiment, an isolated nucleic acid molecule of the invention is at least 15, 20, 25, 30 or more nucleotides in length and hybridizes under stringent conditions to a complement of the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as



Accession Number \_\_\_\_\_. In another embodiment, the nucleic acid is at least 50-100, 100-250, 250-500, 500-750, 750-1000, 1000-1250, 1250-1500, 1500-1700 or more nucleotides in length.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences that are significantly identical or homologous to each other remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% identical to each other remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., John Wiley & Sons, Inc. (1995), sections 2, 4 and 6. Additional stringent conditions can be found in *Molecular Cloning: A Laboratory Manual*, Sambrook *et al.*, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), chapters 7, 9 and 11. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 4X sodium chloride/sodium citrate (SSC), at about 65-70°C (or alternatively hybridization in 4X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 1X SSC, at about 65-70°C. A preferred, non-limiting example of highly stringent hybridization conditions are hybridization in 1X SSC, at about 65-70°C (or alternatively hybridization in 1X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 0.3X SSC, at about 65-70°C. A preferred, non-limiting example of reduced stringency hybridization conditions are hybridization in 4X SSC, at about 50-60°C (or alternatively hybridization in 6X SSC plus 50% formamide at about 40-45°C) followed by one or more washes in 2X SSC, at about 50-60°C. Ranges intermediate to the above-recited values, *e.g.*, at 65-70 °C or at 42-50°C are also intended to be encompassed by the present invention. SSPE (1xSSPE is 0.15M NaCl, 10mM NaH<sub>2</sub>PO<sub>4</sub>, and 1.25mM EDTA, pH 7.4) can be substituted for SSC (1xSSC is 0.15M NaCl and 15mM sodium citrate) in the hybridization and wash buffers; washes are performed for 15 minutes after hybridization is complete. The hybridization temperature for hybrids anticipated to be less than 50 base pairs in length should be 5-10°C less than the melting temperature ( $T_m$ ) of the hybrid, where  $T_m$  is determined according to the following equations. For hybrids less than 18 base pairs in length,  $T_m(^{\circ}\text{C}) = 2(\# \text{ of A} + \text{T bases}) + 4(\# \text{ of G} + \text{C bases})$ . For hybrids between 18 and 49 base pairs in length,  $T_m(^{\circ}\text{C}) = 81.5 + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\% \text{G} + \text{C}) - (600/\text{N})$ , where N is the number of bases in the hybrid, and  $[\text{Na}^+]$  is the concentration of sodium ions in the hybridization buffer ( $[\text{Na}^+]$  for 1xSSC = 0.165 M). It will also be recognized by the skilled practitioner that additional reagents may be added to hybridization and/or wash buffers to decrease non-specific hybridization of nucleic acid molecules to membranes, for example, nitrocellulose or nylon membranes, including but not limited

to blocking agents (*e.g.*, BSA or salmon or herring sperm carrier DNA), detergents (*e.g.*, SDS), chelating agents (*e.g.*, EDTA), Ficoll, PVP and the like. When using nylon membranes, in particular, an additional preferred, non-limiting example of stringent hybridization conditions is hybridization in 0.25-0.5M NaH<sub>2</sub>PO<sub>4</sub>, 7% SDS at about  
5 65°C, followed by one or more washes at 0.02M NaH<sub>2</sub>PO<sub>4</sub>, 1% SDS at 65°C (see *e.g.*, Church and Gilbert (1984) *Proc. Natl. Acad. Sci. USA* 81:1991-1995).

Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a complement of the sequence of SEQ ID NO:1 or 3, and corresponds to a naturally-occurring nucleic acid molecule. As used herein, a  
10 "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In addition to naturally-occurring allelic variants of the 52871 sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:1 or 3, or the  
15 nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, thereby leading to changes in the amino acid sequence of the encoded 52871 proteins, without altering the functional ability of the 52871 proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:1 or 3, or the nucleotide  
20 sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of 52871 (*e.g.*, the sequence of SEQ ID NO:2) without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the 52871 proteins  
25 of the present invention (for example, within transmembrane domains, the 7-TMR profile, or the G-protein coupled receptor signature), are predicted to be particularly unamenable to alteration. Furthermore, additional amino acid residues that are conserved between the 52871 proteins of the present invention and other members of the 52871 family are not likely to be amenable to alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding 52871 proteins that contain changes in amino acid residues that are not essential for activity. Such 52871 proteins differ in amino acid sequence from SEQ ID NO:2, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein  
35 comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:2, *e.g.*, to the entire length of SEQ ID NO:2).

An isolated nucleic acid molecule encoding a 52871 protein identical to the protein of SEQ ID NO:2 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as

5 Accession Number \_\_\_\_\_ such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_ by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid

10 substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*,

15 aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino

20 acid residue in a 52871 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a 52871 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for 52871 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1 or 3, or the

25 nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant 52871 protein can be assayed for the ability to metabolize or catabolize biochemical molecules necessary for energy

30 production or storage, permit intra- or intercellular signaling, metabolize or catabolize metabolically important biomolecules (*e.g.* amino acids, nucleic acids), and to detoxify potentially harmful compounds, or to facilitate the neutralization of these molecules.

In addition to the nucleic acid molecules encoding 52871 proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which

35 are antisense thereto. In an exemplary embodiment, the invention provides an isolated nucleic acid molecule which is antisense to a 52871 nucleic acid molecule (*e.g.*, is antisense to the coding strand of a 52871 nucleic acid molecule). An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid

encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire 52871 coding strand, or to only a portion thereof. In one  
5 embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a 52871. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the coding region of human 52871 corresponds to SEQ ID NO:3). In another embodiment, the antisense nucleic acid molecule is antisense to a  
10 "noncoding region" of the coding strand of a nucleotide sequence encoding 52871. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding 52871 disclosed herein (*e.g.*, SEQ ID NO:3), antisense nucleic acids of the invention can be designed according to the rules  
15 of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of 52871 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of 52871 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of 52871 mRNA (*e.g.*, between the -10 and  
20 +10 regions of the start site of a gene nucleotide sequence). An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized  
25 using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-  
30 fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-  
35 methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-

methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5- oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a 52871 protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *F.E.B.S. Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave 52871 mRNA transcripts to thereby inhibit translation of 52871 mRNA. A ribozyme having specificity for a 52871-encoding nucleic acid can be

designed based upon the nucleotide sequence of a 52871 cDNA disclosed herein (*i.e.*, SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a 52871-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, 52871 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, 52871 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the 52871 (*e.g.*, the 52871 promoter and/or enhancers; *e.g.*, nucleotides 1-107 of SEQ ID NO:1) to form triple helical structures that prevent transcription of the 52871 gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6): 569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

In yet another embodiment, the 52871 nucleic acid molecules of the present invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *et al.* *Proc. Natl. Acad. Sci.* 93: 14670-675.

PNAs of 52871 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of 52871 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (*e.g.*, by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (*e.g.*, S1 nucleases (Hyrup B. (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

In another embodiment, PNAs of 52871 can be modified, (e.g., to enhance their stability or cellular uptake), by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of 52871 nucleic acid molecules can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, (e.g., RNase H and DNA polymerases), to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup B. (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup B. (1996) *supra* and Finn P.J. *et al.* (1996) *Nucleic Acids Res.* 24 (17): 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, e.g., 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. *et al.* (1989) *Nucleic Acid Res.* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn P.J. *et al.* (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5: 1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, e.g., Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents (see, e.g., Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

Alternatively, the expression characteristics of an endogenous 52871 gene within a cell line or microorganism may be modified by inserting a heterologous DNA regulatory element into the genome of a stable cell line or cloned microorganism such that the inserted regulatory element is operatively linked with the endogenous 52871 gene. For example, an endogenous 52871 gene which is normally "transcriptionally silent", i.e., a 52871 gene which is normally not expressed, or is expressed only at very low levels in a cell line or microorganism, may be activated by inserting a regulatory

element which is capable of promoting the expression of a normally expressed gene product in that cell line or microorganism. Alternatively, a transcriptionally silent, endogenous 52871 gene may be activated by insertion of a promiscuous regulatory element that works across cell types.

- 5           A heterologous regulatory element may be inserted into a stable cell line or cloned microorganism, such that it is operatively linked with an endogenous 52871 gene, using techniques, such as targeted homologous recombination, which are well known to those of skill in the art, and described, *e.g.*, in Chappel, U.S. Patent No. 5,272,071; PCT publication No. WO 91/06667, published May 16, 1991.

10

## II. Isolated 52871 Proteins and Anti-52871 Antibodies

- One aspect of the invention pertains to isolated or recombinant 52871 proteins and polypeptides, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-52871 antibodies. In one  
15   embodiment, native 52871 proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, 52871 proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a 52871 protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

- 20           An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the 52871 protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of 52871 protein in which  
25   the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of 52871 protein having less than about 30% (by dry weight) of non-52871 protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-52871 protein, still more preferably less  
30   than about 10% of non-52871 protein, and most preferably less than about 5% non-52871 protein. When the 52871 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

- 35           The language "substantially free of chemical precursors or other chemicals" includes preparations of 52871 protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals"



includes preparations of 52871 protein having less than about 30% (by dry weight) of chemical precursors or non-52871 chemicals, more preferably less than about 20% chemical precursors or non-52871 chemicals, still more preferably less than about 10% chemical precursors or non-52871 chemicals, and most preferably less than about 5% chemical precursors or non-52871 chemicals.

As used herein, a "biologically active portion" of a 52871 protein includes a fragment of a 52871 protein which participates in an interaction between a 52871 molecule and a non-52871 molecule. Biologically active portions of a 52871 protein include peptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the 52871 protein, *e.g.*, the amino acid sequence shown in SEQ ID NO:2, which include less amino acids than the full length 52871 protein, and exhibit at least one activity of a 52871 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the 52871 protein, *e.g.*, signal molecule recognition, signal transduction, modulation of intracellular signaling, modulation of nociception, and/or modulation of pain. A biologically active portion of a 52871 protein can be a polypeptide which is, for example, 25, 50, 75, 100, 125, 150, 175, 200, 250, 300 or more amino acids in length. Biologically active portions of a 52871 protein can be used as targets for developing agents which modulate a 52871 mediated activity, *e.g.*, intercellular signaling, modulation of nociception, and/or modulation of pain.

It is to be understood that a preferred biologically active portion of a 52871 protein of the present invention may contain one or more of the following domains: a transmembrane domain, a 7-TMR profile, a G-protein coupled receptor signature. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native 52871 protein.

Another aspect of the invention features fragments of the protein having the amino acid sequence of SEQ ID NO:2, for example, for use as immunogens. In one embodiment, a fragment comprises at least 5 amino acids (*e.g.*, contiguous or consecutive amino acids) of the amino acid sequence of SEQ ID NO:2, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession Number \_\_\_\_\_. In another embodiment, a fragment comprises at least 10, 15, 20, 25, 30, 35, 40, 45, 50 or more amino acids (*e.g.*, contiguous or consecutive amino acids) of the amino acid sequence of SEQ ID NO:2, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession Number \_\_\_\_\_.

In a preferred embodiment, a 52871 protein has an amino acid sequence shown in SEQ ID NO:2. In other embodiments, the 52871 protein is substantially identical to

SEQ ID NO:2, and retains the functional activity of the protein of SEQ ID NO:2, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. In another embodiment, the 52871 protein is a protein which comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:2.

In another embodiment, the invention features a 52871 protein which is encoded by a nucleic acid molecule consisting of a nucleotide sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to a nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3, or a complement thereof. This invention further features a 52871 protein which is encoded by a nucleic acid molecule consisting of a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3, or a complement thereof.

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-identical sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence (*e.g.*, when aligning a second sequence to the 52871 amino acid sequence of SEQ ID NO:2 having 400 amino acid residues, at least 50, preferably at least 100, more preferably at least 150, even more preferably at least 200, and even more preferably at least 300 or more amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined

using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet  
5 another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A preferred, non-limiting example of parameters to be used in conjunction with the GAP program include a Blossum 62  
10 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of Meyers and Miller (*Comput. Appl. Biosci.*, 4:11-17 (1988)) which has been incorporated into the ALIGN program  
15 (version 2.0 or version 2.U), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed  
20 using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 52871 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 100, wordlength = 3 to obtain  
25 amino acid sequences homologous to 52871 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

30 The invention also provides 52871 chimeric or fusion proteins. As used herein, a 52871 "chimeric protein" or "fusion protein" comprises a 52871 polypeptide operatively linked to a non-52871 polypeptide. An "52871 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a 52871 molecule, whereas a "non-52871 polypeptide" refers to a polypeptide having an amino acid sequence  
35 corresponding to a protein which is not substantially homologous to the 52871 protein, *e.g.*, a protein which is different from the 52871 protein and which is derived from the same or a different organism. Within a 52871 fusion protein the 52871 polypeptide can correspond to all or a portion of a 52871 protein. In a preferred embodiment, a 52871

fusion protein comprises at least one biologically active portion of a 52871 protein. In another preferred embodiment, a 52871 fusion protein comprises at least two biologically active portions of a 52871 protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the 52871 polypeptide and the non-52871 polypeptide are fused in-frame to each other. The non-52871 polypeptide can be fused to the N-terminus or C-terminus of the 52871 polypeptide.

For example, in one embodiment, the fusion protein is a GST-52871 fusion protein in which the 52871 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant 52871.

In another embodiment, the fusion protein is a 52871 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of 52871 can be increased through use of a heterologous signal sequence.

The 52871 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The 52871 fusion proteins can be used to affect the bioavailability of a 52871 substrate. Use of 52871 fusion proteins may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a 52871 protein; (ii) mis-regulation of the 52871 gene; and (iii) aberrant post-translational modification of a 52871 protein.

Moreover, the 52871-fusion proteins of the invention can be used as immunogens to produce anti-52871 antibodies in a subject for use in screening assays to identify molecules which inhibit the interaction of 52871 with a 52871 substrate.

Preferably, a 52871 chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A 52871-

encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the 52871 protein.

The present invention also pertains to variants of the 52871 proteins which function as either 52871 agonists (mimetics) or as 52871 antagonists. Variants of the 52871 proteins can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of a 52871 protein. An agonist of the 52871 proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a 52871 protein. An antagonist of a 52871 protein can inhibit one or more of the activities of the naturally occurring form of the 52871 protein by, for example, competitively modulating a 52871-mediated activity of a 52871 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the 52871 protein.

In one embodiment, variants of a 52871 protein which function as either 52871 agonists (mimetics) or as 52871 antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of a 52871 protein for 52871 protein agonist or antagonist activity. In one embodiment, a variegated library of 52871 variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of 52871 variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential 52871 sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of 52871 sequences therein. There are a variety of methods which can be used to produce libraries of potential 52871 variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential 52871 sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477).

In addition, libraries of fragments of a 52871 protein coding sequence can be used to generate a variegated population of 52871 fragments for screening and subsequent selection of variants of a 52871 protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a 52871 coding sequence with a nuclease under conditions wherein nicking

occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression  
5 vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the 52871 protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for  
10 rapid screening of the gene libraries generated by the combinatorial mutagenesis of 52871 proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which  
15 detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify 52871 variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein*  
20 *Engineering* 6(3): 327-331).

In one embodiment, cell based assays can be exploited to analyze a variegated 52871 library. For example, a library of expression vectors can be transfected into a cell line, *e.g.*, a neuronal cell line, which ordinarily responds to a 52871 ligand in a particular 52871 ligand-dependent manner. The transfected cells are then contacted with a 52871  
25 ligand and the effect of expression of the mutant on, *e.g.*, membrane excitability of 52871 can be detected. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the 52871 ligand, and the individual clones further characterized.

An isolated 52871 protein, or a portion or fragment thereof, can be used as an  
30 immunogen to generate antibodies that bind 52871 using standard techniques for polyclonal and monoclonal antibody preparation. A full-length 52871 protein can be used or, alternatively, the invention provides antigenic peptide fragments of 52871 for use as immunogens. The antigenic peptide of 52871 comprises at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an  
35 epitope of 52871 such that an antibody raised against the peptide forms a specific immune complex with the 52871 protein. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more

preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Preferred epitopes encompassed by the antigenic peptide are regions of 52871 that are located on the surface of the protein, *e.g.*, hydrophilic regions, as well as regions  
5 with high antigenicity.

A 52871 immunogen typically is used to prepare antibodies by immunizing a suitable subject, (*e.g.*, rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed 52871 protein or a chemically synthesized 52871 polypeptide. The  
10 preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic 52871 preparation induces a polyclonal anti-52871 antibody response.

Accordingly, another aspect of the invention pertains to anti-52871 antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and  
15 immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as a 52871. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')<sub>2</sub> fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and  
20 monoclonal antibodies that bind 52871 molecules. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of 52871. A monoclonal antibody composition thus typically displays a single binding affinity for a particular 52871  
25 protein with which it immunoreacts.

Polyclonal anti-52871 antibodies can be prepared as described above by immunizing a suitable subject with a 52871 immunogen. The anti-52871 antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized 52871. If  
30 desired, the antibody molecules directed against 52871 can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the anti-52871 antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard  
35 techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown *et al.* (1981) *J. Immunol.* 127:539-46; Brown *et al.* (1980) *J. Biol. Chem.* 255:4980-83; Yeh *et al.* (1976) *Proc. Natl. Acad. Sci. USA* 76:2927-31; and Yeh *et al.* (1982) *Int. J. Cancer* 29:269-75),

the more recent human B cell hybridoma technique (Kozbor *et al.* (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole *et al.* (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally  
5 R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter *et al.* (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a 52871 immunogen as described above,  
10 and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds 52871.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-52871 monoclonal antibody (see, *e.g.*, G. Galfré *et al.* (1977) *Nature* 266:55052; Gefter *et al.*  
15 *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (*e.g.*, a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made  
20 by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques,  
25 *e.g.*, the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days  
30 because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind 52871, *e.g.*, using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-52871 antibody can be identified and isolated by screening a  
35 recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with 52871 to thereby isolate immunoglobulin library members that bind 52871. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01;



and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner *et al.* U.S. Patent No. 5,223,409; Kang *et al.* PCT International Publication No. WO 92/18619; 5 Dower *et al.* PCT International Publication No. WO 91/17271; Winter *et al.* PCT International Publication WO 92/20791; Markland *et al.* PCT International Publication No. WO 92/15679; Breitling *et al.* PCT International Publication WO 93/01288; McCafferty *et al.* PCT International Publication No. WO 92/01047; Garrard *et al.* PCT International Publication No. WO 92/09690; Ladner *et al.* PCT International 10 Publication No. WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J* 12:725-734; Hawkins *et al.* (1992) *J. Mol. Biol.* 226:889-896; Clarkson *et al.* (1991) *Nature* 352:624-628; Gram *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3576-3580; Garrard *et al.* (1991) 15 *Bio/Technology* 9:1373-1377; Hoogenboom *et al.* (1991) *Nuc. Acid Res.* 19:4133-4137; Barbas *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:7978-7982; and McCafferty *et al.* *Nature* (1990) 348:552-554.

Additionally, recombinant anti-52871 antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, 20 which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson *et al.* International Application No. PCT/US86/02269; Akira, *et al.* European Patent Application 184,187; Taniguchi, M., European Patent Application 25 171,496; Morrison *et al.* European Patent Application 173,494; Neuberger *et al.* PCT International Publication No. WO 86/01533; Cabilly *et al.* U.S. Patent No. 4,816,567; Cabilly *et al.* European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 30 84:214-218; Nishimura *et al.* (1987) *Canc. Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559; Morrison, S. L. (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeven *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

35 An anti-52871 antibody (e.g., monoclonal antibody) can be used to isolate 52871 by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-52871 antibody can facilitate the purification of natural 52871 from cells and of recombinantly produced 52871 expressed in host cells. Moreover, an anti-52871

antibody can be used to detect 52871 protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the 52871 protein. Anti-52871 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy  
5 of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or  
10 acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include  
15 luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

### III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, for example recombinant  
20 expression vectors, containing a 52871 nucleic acid molecule or vectors containing a nucleic acid molecule which encodes a 52871 protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA  
25 segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host  
30 cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can  
35 be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is  
5 operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory  
10 sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cells and  
15 those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce  
20 proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., 52871 proteins, mutant forms of 52871 proteins, fusion proteins, and the like).

Accordingly, an exemplary embodiment provides a method for producing a protein, preferably a 52871 protein, by culturing in a suitable medium a host cell of the  
25 invention (e.g., a mammalian host cell such as a non-human mammalian cell) containing a recombinant expression vector, such that the protein is produced.

The recombinant expression vectors of the invention can be designed for expression of 52871 proteins in prokaryotic or eukaryotic cells. For example, 52871 proteins can be expressed in bacterial cells such as *E. coli*, insect cells (using  
30 baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

35 Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion

vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be utilized in 52871 activity assays, (e.g., direct assays or competitive assays described in detail below), or to generate antibodies specific for 52871 proteins, for example. In a preferred embodiment, a 52871 fusion protein expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six (6) weeks).

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the 52871 expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (InVitrogen Corp, San Diego, CA).

Alternatively, 52871 proteins can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense

orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to 52871 mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a 52871 nucleic acid molecule of the invention is introduced, *e.g.*, a 52871 nucleic acid molecule within a vector (*e.g.*, a recombinant expression vector) or a 52871 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a 52871 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may

integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a 52871 protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) a 52871 protein. Accordingly, the invention further provides methods for producing a 52871 protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of the invention (into which a recombinant expression vector encoding a 52871 protein has been introduced) in a suitable medium such that a 52871 protein is produced. In another embodiment, the method further comprises isolating a 52871 protein from the medium or the host cell.

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which 52871-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous 52871 sequences have been introduced into their genome or homologous recombinant animals in which endogenous 52871 sequences have been altered. Such animals are useful for studying the function and/or activity of a 52871 and for identifying and/or evaluating modulators of 52871 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous 52871 gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing a 52871-encoding nucleic acid into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The 52871 cDNA sequence of SEQ ID NO:1 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of a human 52871 gene, such as a mouse or rat 52871 gene, can be used as a transgene. Alternatively, a 52871 gene homologue, such as another 52871 family member, can be isolated based on hybridization to the 52871 cDNA sequences of SEQ ID NO:1 or 3, or the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_ (described further in subsection I above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a 52871 transgene to direct expression of a 52871 protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of a 52871 transgene in its genome and/or expression of 52871 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a 52871 protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a 52871 gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the 52871 gene. The 52871 gene can be a human gene (*e.g.*, the cDNA of SEQ ID NO:3), but more preferably, is a non-human homologue of a human 52871 gene (*e.g.*, a cDNA isolated by stringent hybridization with the nucleotide sequence of SEQ ID NO:1). For example, a mouse 52871 gene can be used to construct a homologous recombination nucleic acid molecule, *e.g.*, a vector, suitable for altering an endogenous 52871 gene in the mouse genome. In a preferred embodiment, the homologous recombination nucleic acid molecule is designed such that, upon homologous recombination, the endogenous 52871 gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the homologous recombination nucleic acid molecule can be designed such that, upon homologous recombination, the endogenous



52871 gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous 52871 protein). In the homologous recombination nucleic acid molecule, the altered portion of the 52871 gene is flanked at its 5' and 3' ends by additional nucleic acid sequence of the 52871 gene to allow for homologous recombination to occur between the exogenous 52871 gene carried by the homologous recombination nucleic acid molecule and an endogenous 52871 gene in a cell, e.g., an embryonic stem cell. The additional flanking 52871 nucleic acid sequence is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the homologous recombination nucleic acid molecule (see, e.g., Thomas, K.R. and Capecchi, M. R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The homologous recombination nucleic acid molecule is introduced into a cell, e.g., an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced 52871 gene has homologously recombined with the endogenous 52871 gene are selected (see e.g., Li, E. *et al.* (1992) *Cell* 69:915). The selected cells can then be injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination nucleic acid molecules, e.g., vectors, or homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec *et al.*; WO 91/01140 by Smithies *et al.*; WO 92/0968 by Zijlstra *et al.*; and WO 93/04169 by Berns *et al.*

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a

transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al.* (1997) *Nature* 385:810-813 and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, *e.g.*, a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G<sub>0</sub> phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, *e.g.*, the somatic cell, is isolated.

#### IV. Pharmaceutical Compositions

The 52871 nucleic acid molecules, 52871 proteins, fragments thereof, anti-52871 antibodies, and 52871 modulators (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be

enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the  
5 extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage  
10 and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the  
15 maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium  
20 chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a fragment of a 52871 protein or an anti-52871 antibody) in the  
25 required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions,  
30 the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral  
35 therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically

compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required

pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

- 5           Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio
- 10   LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

- The data obtained from the cell culture assays and animal studies can be used in
- 15   formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially
- 20   from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high
- 25   performance liquid chromatography.

- As defined herein, a therapeutically effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to
- 30   7 mg/kg, or 5 to 6 mg/kg body weight. The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can
- 35   include a single treatment or, preferably, can include a series of treatments.

          In a preferred example, a subject is treated with antibody, protein, or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between

about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody, protein, or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as  
5 described herein.

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs,  
10 organic or inorganic compounds (*i.e.*, including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than  
15 about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds. It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being  
20 treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention.

Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (*e.g.*, about 1 microgram per kilogram to about  
25 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one  
30 or more of these small molecules is to be administered to an animal (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any  
35 particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of

excretion, any drug combination, and the degree of expression or activity to be modulated.

Further, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include  
5 taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs  
10 or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II)  
15 (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

The conjugates of the invention can be used for modifying a given biological  
20 response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet  
25 derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well  
30 known, see, *e.g.*, Arnon *et al.*, "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in *Monoclonal Antibodies And Cancer Therapy*, Reisfeld *et al.* (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom *et al.*, "Antibodies For Drug Delivery", in *Controlled Drug Delivery* (2nd Ed.), Robinson *et al.* (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer  
35 Therapy: A Review", in *Monoclonal Antibodies '84: Biological And Clinical Applications*, Pinchera *et al.* (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in *Monoclonal Antibodies For Cancer Detection And Therapy*, Baldwin *et al.* (eds.), pp.

303-16 (Academic Press 1985), and Thorpe *et al.*, "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", *Immunol. Rev.*, 62:119-58 (1982).

Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

5           The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see *e.g.*, Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include  
10 the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

15           The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

#### V. Uses and Methods of the Invention

          The nucleic acid molecules, proteins, protein homologues, modulators, and  
20 antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (*e.g.*, therapeutic and prophylactic). As described herein, a 52871 protein of the invention has one or more of the following activities: (i) interaction of a 52871 protein with soluble  
25 52871 ligand; (ii) interaction of a 52871 protein with a membrane-bound non-52871 protein; (iii) interaction of a 52871 protein with an intracellular protein (*e.g.*, an intracellular enzyme or signal transduction molecule); (iv) indirect interaction of a 52871 protein with an intracellular protein (*e.g.*, a downstream signal transduction molecule); v) modulation of intra- or intercellular signaling or feedback mechanisms; vi)  
30 modulation of the intracellular levels and/or homeostatic balance of signaling molecule pools (*e.g.*,  $\text{Ca}^{2+}$ , diacylglycerol,  $\text{IP}_3$ , cAMP); vii) regulation of cellular proliferation; viii) regulation of cellular differentiation; ix) regulation of development; x) regulation of gene expression in a cell which expresses a 52871 protein; xi) regulation of nociception; xii) regulation of pain signaling; and xiii) regulation of cell death. As such, a 52871  
35 molecule of the invention (or modulator thereof) can be used, for example, in at least one of the following activities: 1) modulation of intra- or intercellular signaling or feedback mechanisms; 2) modulation of the intracellular levels and/or homeostatic balance of signaling molecule pools (*e.g.*,  $\text{Ca}^{2+}$ , diacylglycerol,  $\text{IP}_3$ , cAMP); 3)



modulation of cellular proliferation; 4) modulation of cellular differentiation or development; 5) modulation of gene expression in a cell which expresses a 52871 protein; 6) modulation of nociception; 7) modulation of pain; and 8) modulation of cell death.

5           Modulation of 52871 activity has particular applicability in treating, for example, disorders characterized by insufficient or excessive production of 52871 protein or production of 52871 protein forms which have decreased, or aberrant or unwanted activity compared to 52871 wild type protein, preferably GPCR-associated disorders. Moreover, modulation of 52871 activity has particular application in treating  
10   pain and or pain disorders. Modulation of 52871 activity includes, but is not limited to, increasing or enhancing the activity of 52871 (*e.g.*, increasing or enhancing 52871 signaling), decreasing or inhibiting the activity of 52871 (*e.g.*, decreasing or inhibiting 52871 signaling), regulating 52871 cellular localization, trafficking and/or desensitization of 52871.

15           As used herein, a “GPCR-associated disorder” (or a “52871-associated disorder”) includes a disorder, disease or condition which is caused or characterized by a misregulation (*e.g.*, downregulation or upregulation) of a G protein coupled receptor activity (*e.g.*, GPCR-mediated activity), for example, signal molecule recognition activity or a signal transduction activity. GPCR-associated disorders can detrimentally  
20   affect cellular functions such as cellular proliferation, growth, differentiation, or migration, cellular regulation of homeostasis, inter- or intra-cellular communication; nociception; pain; tissue function, such as cardiac function or musculoskeletal function; systemic responses in an organism, such as nervous system responses, hormonal responses (*e.g.*, insulin response), or immune responses; and protection of cells from  
25   toxic compounds (*e.g.*, carcinogens, toxins, mutagens, and toxic byproducts of metabolic activity (*e.g.*, reactive oxygen species). As used herein, the term “pain disorder” includes disorders characterized by aberrant (*e.g.*, excessive or amplified) pain signaling in addition to symptoms and/or phenotypes which result from wild-type, or normal, pain signaling mechanisms. Examples of pain disorders include posttherapeutic  
30   neuralgia, diabetic neuropathy, postmastectomy pain syndrome, stump pain, reflex sympathetic dystrophy, trigeminal neuralgia, neuropathic pain, orofacial neuropathic pain, osteoarthritis, rheumatoid arthritis, fibromyalgia syndrome, tension myalgia, Guillian-Barre syndrome, Meralgia paraesthetica, burning mouth syndrome, fibrocitis, myofascial pain syndrome, idiopathic pain disorder, temporomandibular joint syndrome,  
35   atypical odontalgia, loin pain, haematuria syndrome, non-cardiac chest pain, low back pain, chronic nonspecific pain, psychogenic pain, musculoskeletal pain disorder, chronic pelvic pain, nonorganic chronic headache, tension-type headache, cluster headache, migraine, complex regional pain syndrome, vaginismus, nerve trunk pain, somatoform

pain disorder, cyclical mastalgia, chronic fatigue syndrome, multiple somatization syndrome, chronic pain disorder, somatization disorder, Syndrome X, facial pain, idiopathic pain disorder, posttraumatic rheumatic pain modulation disorder (fibrositis syndrome), hyperalgesia, and Tangier disease. As used herein, the term "pain signaling mechanisms" includes the cellular mechanisms involved in the development and regulation of pain, *e.g.*, pain elicited by noxious chemical, mechanical, or thermal stimuli, in a subject, *e.g.*, a mammal such as a human. In mammals, the initial detection of noxious chemical, mechanical, or thermal stimuli, a process referred to as "nociception", occurs predominantly at the peripheral terminals of specialized, small diameter sensory neurons. These sensory neurons transmit the information to the central nervous system, evoking a perception of pain or discomfort and initiating appropriate protective reflexes. The 52781 molecules of the present invention are expressed in brain, spinal cord, skin and dorsal root ganglia, and, thus, may be involved in the detection of these noxious chemical, mechanical, or thermal stimuli and/or in the transduction of this information into membrane depolarization events. Thus, the 52781 molecules, by participating in pain signaling mechanisms, may modulate pain elicitation and act as targets for developing novel diagnostic targets and therapeutic agents to control pain and pain disorders.

Further examples of GPCR- or 52871-associated disorders include CNS disorders such as cognitive and neurodegenerative disorders, examples of which include, but are not limited to, Alzheimer's disease, dementias related to Alzheimer's disease (such as Pick's disease), Parkinson's and other Lewy diffuse body diseases, senile dementia, Huntington's disease, Gilles de la Tourette's syndrome, multiple sclerosis, amyotrophic lateral sclerosis, progressive supranuclear palsy, epilepsy, and Jakob-Creutzfeldt disease; autonomic function disorders such as hypertension and sleep disorders, and neuropsychiatric disorders, such as depression, schizophrenia, schizoaffective disorder, korsakoff's psychosis, mania, anxiety disorders, or phobic disorders; learning or memory disorders, *e.g.*, amnesia or age-related memory loss, attention deficit disorder, dysthymic disorder, major depressive disorder, mania, obsessive-compulsive disorder, psychoactive substance use disorders, anxiety, phobias, panic disorder, as well as bipolar affective disorder, *e.g.*, severe bipolar affective (mood) disorder (BP-1), and bipolar affective neurological disorders, *e.g.*, migraine and obesity. Further CNS-related disorders include, for example, those listed in the American Psychiatric Association's Diagnostic and Statistical manual of Mental Disorders (DSM), the most current version of which is incorporated herein by reference in its entirety.

Further examples of GPCR- or 52871-associated disorders include cardiac-related disorders. Cardiovascular system disorders in which the 52871 molecules of the invention may be directly or indirectly involved include arteriosclerosis, ischemia

reperfusion injury, restenosis, arterial inflammation, vascular wall remodeling, ventricular remodeling, rapid ventricular pacing, coronary microembolism, tachycardia, bradycardia, pressure overload, aortic bending, coronary artery ligation, vascular heart disease, atrial fibrillation, Jervell syndrome, Lange-Nielsen syndrome, long-QT  
5 syndrome, congestive heart failure, sinus node dysfunction, angina, heart failure, hypertension, atrial fibrillation, atrial flutter, dilated cardiomyopathy, idiopathic cardiomyopathy, myocardial infarction, coronary artery disease, coronary artery spasm, and arrhythmia. 52871-mediated or related disorders also include disorders of the musculoskeletal system such as paralysis and muscle weakness, *e.g.*, ataxia, myotonia,  
10 and myokymia.

GPCR- or 52871-associated disorders also include cellular proliferation, growth, differentiation, or migration disorders. Cellular proliferation, growth, differentiation, or migration disorders include those disorders that affect cell proliferation, growth, differentiation, or migration processes. As used herein, a “cellular proliferation, growth,  
15 differentiation, or migration process” is a process by which a cell increases in number, size or content, by which a cell develops a specialized set of characteristics which differ from that of other cells, or by which a cell moves closer to or further from a particular location or stimulus. The 52871 molecules of the present invention are involved in signal transduction mechanisms, which are known to be involved in cellular growth,  
20 differentiation, and migration processes. Thus, the 52871 molecules may modulate cellular growth, differentiation, or migration, and may play a role in disorders characterized by aberrantly regulated growth, differentiation, or migration. Such disorders include cancer, *e.g.*, carcinoma, sarcoma, or leukemia; tumor angiogenesis and metastasis; skeletal dysplasia; hepatic disorders; and hematopoietic and/or  
25 myeloproliferative disorders

GPCR- or 52871-associated or related disorders also include hormonal disorders, such as conditions or diseases in which the production and/or regulation of hormones in an organism is aberrant. Examples of such disorders and diseases include type I and type II diabetes mellitus, pituitary disorders (*e.g.*, growth disorders), thyroid disorders  
30 (*e.g.*, hypothyroidism or hyperthyroidism), and reproductive or fertility disorders (*e.g.*, disorders which affect the organs of the reproductive system, *e.g.*, the prostate gland, the uterus, or the vagina; disorders which involve an imbalance in the levels of a reproductive hormone in a subject; disorders affecting the ability of a subject to reproduce; and disorders affecting secondary sex characteristic development, *e.g.*,  
35 adrenal hyperplasia).

52871-associated or related disorders also include immune disorders, such as autoimmune disorders or immune deficiency disorders, *e.g.*, congenital X-linked infantile hypogammaglobulinemia, transient hypogammaglobulinemia, common

variable immunodeficiency, selective IgA deficiency, chronic mucocutaneous candidiasis, or severe combined immunodeficiency. 52871-associated or related disorders also include disorders affecting tissues in which 52871 protein is expressed (e.g., brain, spinal cord, dorsal root ganglia, or skin tissue).

5 In addition, 52871-associated or related disorders also include disorders such as retinitis pigmentosa, stationary night blindness, color blindness, isolated glucocorticoid deficiency, hyperfunctioning thyroid adenoma, familial precocious puberty, familial hypocalciuric hypercalcemia, neonatal severe hyperparathyroidism, and nephrogenic diabetes insipidus.

10 The isolated nucleic acid molecules of the invention can be used, for example, to express 52871 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect 52871 mRNA (e.g., in a biological sample) or a genetic alteration in a 52871 gene, and to modulate 52871 activity, as described further below. The 52871 proteins can be used to treat disorders characterized by insufficient or  
15 excessive production of a 52871 protein and/or 52871 ligand. In addition, the 52871 proteins can be used to screen drugs or compounds which modulate the 52871 activity as well as to treat disorders characterized by insufficient or excessive production of 52871 protein or production of 52871 protein forms which have decreased or aberrant activity compared to 52871 wild type protein. Moreover, the anti-52871 antibodies of  
20 the invention can be used to detect and isolate 52871 proteins, regulate the bioavailability of 52871 proteins, and modulate 52871 activity.

#### A. Screening Assays:

The invention provides a method (also referred to herein as a "screening assay")  
25 for identifying modulators, i.e., candidate or test compounds or agents (e.g., peptides, peptidomimetics, small molecules or other drugs) which bind to 52871 proteins, or have a stimulatory or inhibitory effect on, for example, 52871 expression or 52871 activity.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a 52871 protein or polypeptide or  
30 biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity  
35 chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. 1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a 52871 protein on the cell surface is contacted with a test compound and the ability of the test compound to bind to the 52871 protein determined. The cell, for example, can be of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to a 52871 protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the 52871 protein can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^3\text{H}$ , either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a test compound to interact with a 52871 protein without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a test compound with a 52871 protein without the labeling of either the test compound or the receptor. McConnell *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (*e.g.*, Cytosensor<sup>TM</sup>) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between ligand and receptor.

In a preferred embodiment, the assay comprises contacting a cell which expresses a 52871 protein or biologically active portion thereof, on the cell surface with a 52871 ligand (*e.g.*, a peptide, a neurotransmitter, or a hormone), to form an assay

mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the 52871 protein or biologically active portion thereof, wherein determining the ability of the test compound to interact with the 52871 protein or biologically active portion thereof, comprises determining the ability of the test compound to preferentially bind to the 52871 protein or biologically active portion thereof, as compared to the ability of the 52871 ligand to bind to the 52871 protein or biologically active portion thereof. Determining the ability of the 52871 ligand or 52871 modulator to bind to or interact with a 52871 protein or biologically active portion thereof, can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the 52871 ligand or modulator to bind to or interact with a 52871 protein or biologically active portion thereof, can be accomplished by determining the activity of a 52871 protein or of a downstream 52871 target molecule. For example, the target molecule can be a cellular second messenger, and the activity of the target molecule can be determined by detecting induction of the target (*i.e.* intracellular  $\text{Ca}^{2+}$ , diacylglycerol,  $\text{IP}_3$ , etc.), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (comprising a 52871-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, a proliferative response, a differentiation response, or a signaling response. Accordingly, in one embodiment, the present invention involves a method of identifying a compound which modulates the activity of a 52871 protein, comprising contacting a cell which expresses a 52871 protein with a test compound, determining the ability of the test compound to modulate the activity the 52871 protein, and identifying the compound as a modulator of 52871 activity. In another embodiment, the present invention involves a method of identifying a compound which modulates the activity of a 52871 protein, comprising contacting a cell which expresses a 52871 protein with a test compound, determining the ability of the test compound to modulate the activity of a downstream 52871 target molecule, and identifying the compound as a modulator of 52871 activity.

In yet another embodiment, an assay of the present invention is a cell-free assay in which a 52871 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the 52871 protein or biologically active portion thereof is determined. Binding of the test compound to the 52871 protein can be determined either directly or indirectly as described above.

Binding of the test compound to the 52871 protein can also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 3:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for

studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore™). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In a preferred embodiment, the assay includes contacting the 52871 protein or  
5 biologically active portion thereof with a known ligand (e.g., a peptide, a neurotransmitter, or a hormone) which binds 52871 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a 52871 protein, wherein determining the ability of the test compound to interact with a 52871 protein comprises determining the ability of the test  
10 compound to preferentially bind to 52871 or biologically active portion thereof as compared to the known ligand.

In another embodiment, the assay is a cell-free assay in which a 52871 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the 52871  
15 protein or biologically active portion thereof is determined. Determining the ability of the test compound to modulate the activity of a 52871 protein can be accomplished, for example, by determining the ability of the 52871 protein to modulate the activity of a downstream 52871 target molecule by one of the methods described above for cell-based assays. For example, the catalytic/enzymatic activity of the target molecule on an  
20 appropriate substrate can be determined as previously described. In yet another embodiment, the cell-free assay involves contacting a 52871 protein or biologically active portion thereof with a known ligand (e.g., a peptide, a neurotransmitter, or a hormone) which binds the 52871 protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to  
25 interact with the 52871 protein, wherein determining the ability of the test compound to interact with the 52871 protein comprises determining the ability of the test compound to preferentially bind to or modulate the activity of a 52871 target molecule, as compared to the known ligand.

The cell-free assays of the present invention are amenable to use of both soluble  
30 and/or membrane-bound forms of isolated proteins (e.g. 52871 proteins or biologically active portions thereof or 52871 proteins). In the case of cell-free assays in which a membrane-bound form an isolated protein is used (e.g., a 52871 protein) it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the isolated protein is maintained in solution. Examples of such solubilizing agents include  
35 non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)<sub>n</sub>, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-

cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either 52871 or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a 52871 protein, or interaction of a 52871 protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/52871 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or 52871 protein, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of 52871 binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a 52871 protein or a 52871 target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated 52871 protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with 52871 protein or target molecules but which do not interfere with binding of the 52871 protein to its target molecule can be derivatized to the wells of the plate, and unbound target or 52871 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the 52871 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the 52871 protein or target molecule.

In another embodiment, modulators of 52871 expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of



52871 mRNA or protein in the cell is determined. The level of expression of 52871 mRNA or protein in the presence of the candidate compound is compared to the level of expression of 52871 mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of 52871 expression based on this comparison. For example, when expression of 52871 mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of 52871 mRNA or protein expression. Alternatively, when expression of 52871 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of 52871 mRNA or protein expression. The level of 52871 mRNA or protein expression in the cells can be determined by methods described herein for detecting 52871 mRNA or protein.

In yet another aspect of the invention, the 52871 proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300) to identify other proteins, which bind to or interact with 52871 ("52871-binding proteins" or "52871-bp") and are involved in 52871 activity. Such 52871-binding proteins are also likely to be involved in the propagation of signals by the 52871 proteins as, for example, downstream elements of a 52871-mediated signaling pathway. Alternatively, such 52871-binding proteins are likely to be cell-surface molecules associated with non-52871 expressing cells, wherein such 52871-binding proteins are involved in chemoattraction.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a 52871 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a 52871-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the 52871 protein.

The present invention further features assays (e.g., secondary screening assays or validation assays) designed to confirm the activity of a test compound, for example, as a 52871 modulator. In one embodiment, the invention features screening assays (e.g., secondary screening assays or validation assays) which include administering a test  
5 compound, for example, a test compound that demonstrates binding to a 52871 protein or modulation of a 52871 activity in at least one of the above-described cell-based or cell-free assays, to an animal and determining the ability of the test compound to modulate 52871 activity *in vivo*. Determining the ability of a compound to modulate activity *in vivo* can include, for example, determining the ability of the compound to  
10 modulate signaling activity. Exemplary animals for determining 52871 modulatory activity include normal animals as well as animal models which have one or more signaling dysfunctions. It is also within the scope of this invention to use an agent or compound as described herein (e.g., a 52871 modulating agent, an antisense 52871 nucleic acid molecule, a 52871-specific antibody, or a 52871-binding partner) in an  
15 animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent.

Models for studying pain *in vivo* include rat models of neuropathic pain caused by methods such as intraperitoneal administration of Taxol (Authier *et al.* (2000) *Brain Res.* 887:239-249), chronic constriction injury (CCI), partial sciatic nerve transection (Linenlaub and Sommer (2000) *Pain* 89:97-106), transection of the tibial and sural nerves (Lee *et al.* (2000) *Neurosci. Lett.* 291:29-32), the spared nerve injury model (Decosterd and Woolf (2000) *Pain* 87:149-158), cuffing the sciatic nerve (Pitcher and Henry (2000) *Eur. J. Neurosci.* 12:2006-2020), unilateral tight ligation (Esser and  
25 Sawynok (2000) *Eur. J. Pharmacol.* 399:131-139), L5 spinal nerve ligation (Honroe *et al.* (2000) *Neurosci.* 98:585-598), and photochemically induced ischemic nerve injury (Hao *et al.* (2000) *Exp. Neurol.* 163:231-238); rat models of nociceptive pain caused by methods such as the Chung Method, the Bennett Method, and intraperitoneal administration of complete Freund's adjuvant (CFA) (Abdi *et al.* (2000) *Anesth. Analg.*  
30 91:955-959); rat models of post-incisional pain caused by incising the skin and fascia of a hind paw (Olivera and Prado (2000) *Braz. J. Med. Biol. Res.* 33:957-960); rat models of cancer pain caused by methods such as injecting osteolytic sarcoma cells into the femur (Honroe *et al.* (2000) *Neurosci.* 98:585-598); and rat models of visceral pain caused by methods such as intraperitoneal administration of cyclophosphamide.

35 Various methods of determining an animal's response to pain are known in the art. Examples of such methods include, but are not limited to brief intense exposure to a focused heat source, administration of a noxious chemical subcutaneously, the tail flick test, the hot plate test, the formalin test, Von Frey threshold, and testing for stress-

induced analgesia (*et al.*, by restraint, foot shock, and/or cold water swim) (Crawley (2000) *What's Wrong With My Mouse?* Wiley-Liss pp. 72-75).

This invention further pertains to novel agents identified by the above-described screening assays. Furthermore, this invention pertains to uses of novel agents identified  
5 by the above-described screening assays for treatments as described herein.

## B. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as  
10 polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

15

### 1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the 52871  
20 nucleotide sequences, described herein, can be used to map the location of the 52871 genes on a chromosome. The mapping of the 52871 sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, 52871 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the 52871 nucleotide sequences. Computer  
25 analysis of the 52871 sequences can be used to predict primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the 52871 sequences will yield an amplified fragment.

30 Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains  
35 the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human

chromosomes. (D'Eustachio P. *et al.* (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

5 PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the 52871 nucleotide sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a 52871 sequence to its chromosome include *in situ* hybridization  
10 (described in Fan, Y. *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in  
15 one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence  
20 as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.*, Human Chromosomes: A Manual of Basic Techniques  
25 (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding  
30 sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance  
35 in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland *et al.* (1987) *Nature* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the 52871 gene can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease.

5 Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

10

## 2. Tissue Typing

The 52871 sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for

15 identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers

20 for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the 52871 nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the

25 sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the

30 present invention can be used to obtain such identification sequences from individuals and from tissue. The 52871 nucleotide sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of

35 about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The

noncoding sequences of SEQ ID NO:1 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3 are used, a more appropriate number of primers for positive individual  
5 identification would be 500-2,000.

If a panel of reagents from 52871 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from  
10 extremely small tissue samples.

### 3. Use of 52871 Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator  
15 of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the  
20 origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular  
25 individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique.  
30 Examples of polynucleotide reagents include the 52871 nucleotide sequences or portions thereof, *e.g.*, fragments derived from the noncoding regions of SEQ ID NO:1 having a length of at least 20 bases, preferably at least 30 bases.

The 52871 nucleotide sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or labelable probes which can be used in, for  
35 example, an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, thymus or brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such 52871 probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, *e.g.*, 52871 primers or probes can be used to screen tissue culture for contamination (*i.e.* screen for the presence of a mixture of different types of cells in a culture).

5           C.     Predictive Medicine:

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for  
10 determining 52871 protein and/or nucleic acid expression as well as 52871 activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant or unwanted 52871 expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an  
15 individual is at risk of developing a disorder associated with 52871 protein, nucleic acid expression or activity. For example, mutations in a 52871 gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with 52871 protein, nucleic acid expression or activity.  
20           Another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of 52871 in clinical trials.

These and other agents are described in further detail in the following sections.

25

1.     Diagnostic Assays

An exemplary method for detecting the presence or absence of 52871 protein, polypeptide or nucleic acid in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an  
30 agent capable of detecting 52871 protein, polypeptide or nucleic acid (*e.g.*, mRNA, or genomic DNA) that encodes 52871 protein such that the presence of 52871 protein, polypeptide or nucleic acid is detected in the biological sample. In another aspect, the present invention provides a method for detecting the presence of 52871 activity in a biological sample by contacting the biological sample with an agent capable of detecting  
35 an indicator of 52871 activity such that the presence of 52871 activity is detected in the biological sample. A preferred agent for detecting 52871 mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to 52871 mRNA or genomic DNA. The nucleic acid probe can be, for example, the 52871 nucleic acid set forth in SEQ ID

NO:1 or 3, or the DNA insert of the plasmid deposited with ATCC as Accession Number \_\_\_\_\_, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to 52871 mRNA or genomic DNA or complements thereof. Other  
5 suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting 52871 protein is an antibody capable of binding to 52871 protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')<sub>2</sub>) can be used. The term "labeled", with regard to the probe or  
10 antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with  
15 biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect 52871 mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro*  
20 techniques for detection of 52871 mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of 52871 protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of 52871 genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of 52871 protein  
25 include introducing into a subject a labeled anti-52871 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

The present invention also provides diagnostic assays for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant  
30 modification or mutation of a gene encoding a 52871 protein; (ii) aberrant expression of a gene encoding a 52871 protein; (iii) mis-regulation of the gene; and (iii) aberrant post-translational modification of a 52871 protein, wherein a wild-type form of the gene encodes a protein with a 52871 activity. "Misexpression or aberrant expression", as used herein, refers to a non-wild type pattern of gene expression, at the RNA or protein  
35 level. It includes, but is not limited to, expression at non-wild type levels (e.g., over or under expression); a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed (e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage); a pattern



of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-translational modification, or biological activity of the expressed polypeptide; a pattern of  
5 expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene (e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus).

In one embodiment, the biological sample contains protein molecules from the  
10 test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control  
15 biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting 52871 protein, mRNA, or genomic DNA, such that the presence of 52871 protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of 52871 protein, mRNA or genomic DNA in the control sample with the presence of 52871 protein, mRNA or genomic DNA in the test sample.

20 The invention also encompasses kits for detecting the presence of 52871 in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting 52871 protein or mRNA in a biological sample; means for determining the amount of 52871 in the sample; and means for comparing the amount of 52871 in the sample with a standard. The compound or agent can be packaged in a  
25 suitable container. The kit can further comprise instructions for using the kit to detect 52871 protein or nucleic acid.

## 2. Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify  
30 subjects having or at risk of developing a disease or disorder associated with aberrant or unwanted 52871 expression or activity. As used herein, the term "aberrant" includes a 52871 expression or activity which deviates from the wild type 52871 expression or activity. Aberrant expression or activity includes increased or decreased expression or activity, as well as expression or activity which does not follow the wild type  
35 developmental pattern of expression or the subcellular pattern of expression. For example, aberrant 52871 expression or activity is intended to include the cases in which a mutation in the 52871 gene causes the 52871 gene to be under-expressed or over-expressed and situations in which such mutations result in a non-functional 52871

protein or a protein which does not function in a wild-type fashion, *e.g.*, a protein which does not interact with a 52871 substrate, or one which interacts with a non-52871 substrate. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such as cellular proliferation. For example, the term  
5 unwanted includes a 52871 expression or activity which is undesirable in a subject.

The assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with a misregulation in 52871 protein activity or nucleic acid expression, such as a CNS disorder (*e.g.*, a cognitive or neurodegenerative disorder), a  
10 pain disorder, a cellular proliferation, growth, differentiation, or migration disorder, a cardiovascular disorder, musculoskeletal disorder, an immune disorder, or a hormonal disorder. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disorder associated with a misregulation in 52871 protein activity or nucleic acid expression, such as a CNS disorder, a pain disorder, a cellular  
15 proliferation, growth, differentiation, or migration disorder, a musculoskeletal disorder, a cardiovascular disorder, an immune disorder, or a hormonal disorder. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant or unwanted 52871 expression or activity in which a test sample is obtained from a subject and 52871 protein or nucleic acid (*e.g.*, mRNA or genomic DNA) is  
20 detected, wherein the presence of 52871 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted 52871 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, cerebrospinal fluid or serum), cell sample, or tissue.

25 Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted 52871 expression or activity. For example, such methods can be used to determine whether a subject can be  
30 effectively treated with an agent for a CNS disorder, a pain disorder, a muscular disorder, a cellular proliferation, growth, differentiation, or migration disorder, an immune disorder, or a hormonal disorder. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant or unwanted 52871 expression or activity in which a test  
35 sample is obtained and 52871 protein or nucleic acid expression or activity is detected (*e.g.*, wherein the abundance of 52871 protein or nucleic acid expression or activity is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant or unwanted 52871 expression or activity).

The methods of the invention can also be used to detect genetic alterations in a 52871 gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in 52871 protein activity or nucleic acid expression, such as a CNS disorder, a pain disorder, a musculoskeletal disorder, a cellular proliferation, growth, differentiation, or migration disorder, a cardiovascular disorder, an immune disorder, or a hormonal disorder. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a 52871-protein, or the mis-expression of the 52871 gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a 52871 gene; 2) an addition of one or more nucleotides to a 52871 gene; 3) a substitution of one or more nucleotides of a 52871 gene, 4) a chromosomal rearrangement of a 52871 gene; 5) an alteration in the level of a messenger RNA transcript of a 52871 gene, 6) aberrant modification of a 52871 gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a 52871 gene, 8) a non-wild type level of a 52871-protein, 9) allelic loss of a 52871 gene, and 10) inappropriate post-translational modification of a 52871-protein. As described herein, there are a large number of assays known in the art which can be used for detecting alterations in a 52871 gene. A preferred biological sample is a tissue or serum sample isolated by conventional means from a subject.

In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, *e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, *e.g.*, Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in a 52871 gene (see Abravaya *et al.* (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a 52871 gene under conditions such that hybridization and amplification of the 52871 gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878),

transcriptional amplification system (Kwoh, D.Y. *et al.*, (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. *et al.* (1988) *Bio-Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These  
5 detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a 52871 gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more  
10 restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage  
15 site.

In other embodiments, genetic mutations in 52871 can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Human Mutation* 7: 244-255; Kozal, M.J. *et al.* (1996) *Nature Medicine* 2: 753-  
20 759). For example, genetic mutations in 52871 can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the  
25 identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the 52871 gene and detect mutations by comparing the sequence of the sample 52871 with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxam and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger  
35 ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen *et al.*

(1996) *Adv. Chromatogr.* 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the 52871 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type 52871 sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al.* (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in 52871 cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a 52871 sequence, *e.g.*, a wild-type 52871 sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in 52871 genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control 52871 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting

alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a 52871 gene.

Furthermore, any cell type or tissue in which 52871 is expressed may be utilized in the prognostic assays described herein.

### 3. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs) on the expression or activity of a 52871 protein (*e.g.*, the maintenance of cellular homeostasis) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase 52871 gene expression, protein levels, or upregulate 52871 activity, can be monitored in clinical trials of subjects exhibiting decreased 52871 gene expression, protein levels, or downregulated 52871 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease 52871 gene expression, protein levels, or downregulate 52871 activity, can be monitored in clinical trials of subjects exhibiting increased 52871 gene expression, protein levels, or upregulated 52871 activity. In such clinical trials, the expression or activity of a 52871 gene, and preferably, other genes that have been implicated in, for example, a 52871-associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

For example, and not by way of limitation, genes, including 52871, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) which modulates 52871 activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on 52871-associated disorders (*e.g.*, disorders characterized by deregulated cell proliferation and/or migration or pain disorders), for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of 52871 and other genes implicated in the 52871-associated disorder, respectively. The levels of gene expression (*e.g.*, a gene expression pattern) can be quantified by northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of 52871 or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) including the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a 52871 protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the 52871 protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the 52871 protein, mRNA, or genomic DNA in the pre-administration sample with the 52871 protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of 52871 to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of 52871 to lower levels than detected, *i.e.* to decrease the effectiveness of the agent. According to such an embodiment, 52871 expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

#### D. Methods of Treatment:

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted 52871 expression or activity, *e.g.*, a GPCR-associated disorder, a cell-signaling disorder, pain, or a pain disorder. "Treatment", or "treating" as used herein, is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease or disorder, a symptom of disease or disorder or a predisposition toward a disease or disorder, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease or disorder, the symptoms of the disease or disorder, or the predisposition toward disease. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides. With regard to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes



determine his or her response to a drug (e.g., a patient's "drug response phenotype", or "drug response genotype"). Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the 52871 molecules of the present invention or 52871 modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

#### 1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted 52871 expression or activity, by administering to the subject a 52871 or an agent which modulates 52871 expression or at least one 52871 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted 52871 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the 52871 aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of 52871 aberrancy, for example, a 52871, 52871 agonist or 52871 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

#### 2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating 52871 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell capable of expressing 52871 with an agent that modulates one or more of the activities of 52871 protein activity associated with the cell, such that 52871 activity in the cell is modulated. An agent that modulates 52871 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring substrate molecule of a 52871 protein (e.g., energy transduction metabolites, urea cycle metabolites, lipid metabolism metabolites, amino acid precursors, nucleic acid precursors), a 52871 antibody, a 52871 agonist or antagonist, a peptidomimetic of a 52871 agonist or antagonist, or other small molecule. In one embodiment, the agent stimulates one or more 52871 activities. Examples of such stimulatory agents include active 52871 protein and a nucleic acid molecule encoding 52871 that has been introduced into the cell. In another embodiment, the agent inhibits one or more 52871 activities. Examples of such

inhibitory agents include antisense 52871 nucleic acid molecules, anti-52871 antibodies, and 52871 inhibitors. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a 52871 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) 52871 expression or activity. In another embodiment, the method involves administering a 52871 protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted 52871 expression or activity.

Stimulation of 52871 activity is desirable in situations in which 52871 is abnormally downregulated and/or in which increased 52871 activity is likely to have a beneficial effect. Likewise, inhibition of 52871 activity is desirable in situations in which 52871 is abnormally upregulated and/or in which decreased 52871 activity is likely to have a beneficial effect.

### 3. Pharmacogenomics

The 52871 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on 52871 activity (e.g., 52871 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) 52871-associated disorders (e.g., proliferative disorders, CNS disorders, pain or pain disorders, cardiac disorders, metabolic disorders, or muscular disorders) associated with aberrant or unwanted 52871 activity. In conjunction with such treatment, pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a 52871 molecule or 52871 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a 52871 molecule or 52871 modulator.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. *et al.* (1996) *Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 and Linder, M.W. *et al.* (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic

conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, 5 glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

One pharmacogenomics approach to identifying genes that predict drug 10 response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (*e.g.*, a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically 15 significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of 20 DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of 25 genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug target is known (*e.g.*, a 52871 protein of the present invention), all 30 common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of 35 genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug.

These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (*e.g.*, a 52871 molecule or 52871 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a 52871 molecule or 52871 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

#### 4. Use of 52871 Molecules as Surrogate Markers

The 52871 molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the 52871 molecules of the invention may be detected, and may be correlated with one or more biological states *in vivo*. For example, the 52871 molecules of the invention may serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a "surrogate marker" is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (*e.g.*, with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers may serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of

particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (e.g., early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (e.g., an assessment of cardiovascular disease may be made using cholesterol  
5 levels as a surrogate marker, and an analysis of HIV infection may be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers in the art include: Koomen *et al.* (2000) *J. Mass. Spectrom.* 35: 258-264; and James (1994) *AIDS Treatment News Archive* 209.

10       The 52871 molecules of the invention are also useful as pharmacodynamic markers. As used herein, a "pharmacodynamic marker" is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is  
15 indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker may be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the drug. In this fashion, the distribution or uptake of the drug may be monitored by the pharmacodynamic marker.  
20 Similarly, the presence or quantity of the pharmacodynamic marker may be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug *in vivo*. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a  
25 small amount of a drug may be sufficient to activate multiple rounds of marker (e.g., a 52871 marker) transcription or expression, the amplified marker may be in a quantity which is more readily detectable than the drug itself. Also, the marker may be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-52871 antibodies may be employed in an immune-based detection  
30 system for a 52871 protein marker, or 52871-specific radiolabeled probes may be used to detect a 52871 mRNA marker. Furthermore, the use of a pharmacodynamic marker may offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda *et al.* US 6,033,862; Hattis *et al.* (1991) *Env. Health Perspect.* 90:  
35 229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S21-S24; and Nicolau (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S16-S20.

The 52871 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical

marker which correlates with a specific clinical drug response or susceptibility in a subject (see, e.g., McLeod *et al.* (1999) *Eur. J. Cancer* 35(12): 1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, may be selected. For example, based on the presence or quantity of RNA, or protein (*e.g.*, 52871 protein or RNA) for specific tumor markers in a subject, a drug or course of treatment may be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in 52871 DNA may correlate 52871 drug response. The use of pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

15

#### VI. Electronic Apparatus Readable Media and Arrays

Electronic apparatus readable media comprising 52871 sequence information is also provided. As used herein, "52871 sequence information" refers to any nucleotide and/or amino acid sequence information particular to the 52871 molecules of the present invention, including but not limited to full-length nucleotide and/or amino acid sequences, partial nucleotide and/or amino acid sequences, polymorphic sequences including single nucleotide polymorphisms (SNPs), epitope sequences, and the like. Moreover, information "related to" said 52871 sequence information includes detection of the presence or absence of a sequence (*e.g.*, detection of expression of a sequence, fragment, polymorphism, etc.), determination of the level of a sequence (*e.g.*, detection of a level of expression, for example, a quantitative detection), detection of a reactivity to a sequence (*e.g.*, detection of protein expression and/or levels, for example, using a sequence-specific antibody), and the like. As used herein, "electronic apparatus readable media" refers to any suitable medium for storing, holding or containing data or information that can be read and accessed directly by an electronic apparatus. Such media can include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as compact disc; electronic storage media such as RAM, ROM, EPROM, EEPROM and the like; general hard disks and hybrids of these categories such as magnetic/optical storage media. The medium is adapted or configured for having recorded thereon 52871 sequence information of the present invention.

As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus or other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatus; networks, including a local  
5 area network (LAN), a wide area network (WAN) Internet, Intranet, and Extranet; electronic appliances such as a personal digital assistants (PDAs), cellular phone, pager and the like; and local and distributed processing systems.

As used herein, "recorded" refers to a process for storing or encoding information on the electronic apparatus readable medium. Those skilled in the art can  
10 readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising the 52871 sequence information.

A variety of software programs and formats can be used to store the sequence information on the electronic apparatus readable medium. For example, the sequence information can be represented in a word processing text file, formatted in  
15 commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like, as well as in other forms. Any number of data processor structuring formats (*e.g.*, text file or database) may be employed in order to obtain or create a medium having recorded thereon the 52871 sequence information.

20 By providing 52871 sequence information in readable form, one can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the sequence information in readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequences of the  
25 invention which match a particular target sequence or target motif.

The present invention therefore provides a medium for holding instructions for performing a method for determining whether a subject has a 52871- associated disease or disorder or a pre-disposition to a 52871-associated disease or disorder, wherein the method comprises the steps of determining 52871 sequence information associated with  
30 the subject and based on the 52871 sequence information, determining whether the subject has a 52871 -associated disease or disorder or a pre-disposition to a 52871-associated disease or disorder and/or recommending a particular treatment for the disease, disorder or pre-disease condition.

The present invention further provides in an electronic system and/or in a  
35 network, a method for determining whether a subject has a 52871-associated disease or disorder or a pre-disposition to a disease associated with a 52871 wherein the method comprises the steps of determining 52871 sequence information associated with the subject, and based on the 52871 sequence information, determining whether the subject

has a 52871 -associated disease or disorder or a pre-disposition to a 52871-associated disease or disorder, and/or recommending a particular treatment for the disease, disorder or pre-disease condition. The method may further comprise the step of receiving phenotypic information associated with the subject and/or acquiring from a network  
5 phenotypic information associated with the subject.

The present invention also provides in a network, a method for determining whether a subject has a 52871-associated disease or disorder or a pre-disposition to a 52871 associated disease or disorder associated with 52871, said method comprising the steps of receiving 52871 sequence information from the subject and/or information  
10 related thereto, receiving phenotypic information associated with the subject, acquiring information from the network corresponding to 52871 and/or a 52871-associated disease or disorder, and based on one or more of the phenotypic information, the 52871 information (*e.g.*, sequence information and/or information related thereto), and the acquired information, determining whether the subject has a 52871-associated disease or  
15 disorder or a pre-disposition to a 52871-associated disease or disorder (*e.g.*, a pain disorder). The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

The present invention also provides a business method for determining whether a subject has a 52871-associated disease or disorder or a pre-disposition to a 52871-  
20 associated disease or disorder, said method comprising the steps of receiving information related to 52871 (*e.g.*, sequence information and/or information related thereto), receiving phenotypic information associated with the subject, acquiring information from the network related to 52871 and/or related to a 52871-associated disease or disorder, and based on one or more of the phenotypic information, the 52871  
25 information, and the acquired information, determining whether the subject has a 52871-associated disease or disorder or a pre-disposition to a 52871-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

The invention also includes an array comprising a 52871 sequence of the present  
30 invention. The array can be used to assay expression of one or more genes in the array. In one embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array. In this manner, up to about 7600 genes can be simultaneously assayed for expression, one of which can be 52871. This allows a profile to be developed showing a battery of genes specifically expressed in one or more  
35 tissues.

In addition to such qualitative determination, the invention allows the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue is ascertainable. Thus, genes can be



grouped on the basis of their tissue expression *per se* and level of expression in that tissue. This is useful, for example, in ascertaining the relationship of gene expression between or among tissues. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell  
5 type on another cell type in response to a biological stimulus can be determined. Such a determination is useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the  
10 opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

In another embodiment, the array can be used to monitor the time course of  
15 expression of one or more genes in the array. This can occur in various biological contexts, as disclosed herein, for example development of a 52871-associated disease or disorder, progression of 52871-associated disease or disorder, and processes, such a cellular transformation associated with the 52871-associated disease or disorder.

The array is also useful for ascertaining the effect of the expression of a gene on  
20 the expression of other genes in the same cell or in different cells (*e.g.*, ascertaining the effect of 52871 expression on the expression of other genes). This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

The array is also useful for ascertaining differential expression patterns of one or  
25 more genes in normal and abnormal cells. This provides a battery of genes (*e.g.*, including 52871) that could serve as a molecular target for diagnosis or therapeutic intervention.

This invention is further illustrated by the following examples which should not  
be construed as limiting. The contents of all references, patents and published patent  
30 applications cited throughout this application, as well as the Figures, are incorporated herein by reference.

## EXAMPLES

### 5 EXAMPLE 1: IDENTIFICATION AND CHARACTERIZATION OF HUMAN

#### 52871 cDNA

In this example, the identification and characterization of the gene encoding human 52871 (clone Fbh52871) is described.

10

#### Isolation of the 52871 cDNA

The invention is based, at least in part, on the discovery of a human gene encoding a novel GPCR protein, referred to herein as 52871. The entire sequence of human clone Fbh52871, was determined and found to contain an open reading frame  
15 termed human "52871", set forth in Figure 1. The amino acid sequence of the human 52871 expression product is set forth in Figure 1. The 52871 protein sequence set forth in SEQ ID NO:2 comprises about 348 amino acids and is shown in Figure 1. The coding region (open reading frame) of SEQ ID NO:1, is set forth as SEQ ID NO:3. Clone Fbh52871, comprising the coding region of human 52871, was deposited with the  
20 American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, U.S.A., on \_\_\_\_\_, and assigned Accession No. \_\_\_\_\_.

#### Analysis of the Human 52871 Molecule

25 An analysis of the possible cellular localization of the 52871 protein based on its amino acid sequence was performed using the methods and algorithms described in Nakai and Kanehisa (1992) *Genomics* 14:897-911, and at <http://psort.nibb.ac.jp>. The results of the analysis predict that human 52871 (SEQ ID NO:2) is localized intracellularly (probabilities are shown for localization to, *e.g.*, 44.4% in the  
30 endoplasmic reticulum, 22.2% in the vacuoles, 11.1% in the golgi apparatus, 11.1% in the vesicles of the secretory system, and 11.1% in the mitochondria).

A search of the amino acid sequence of 52871 was performed against the Memsat database. This search resulted in the identification of seven transmembrane domains in the amino acid sequence of human 52871 (SEQ ID NO:2) at about residues  
35 53-75, 90-108, 126-144, 165-186, 210-234, 275-293, and 309-333.

Examination of the amino acid sequence of 52871 (SEQ ID NO:2) revealed that the present invention contains conserved cysteines found in the first 2 extracellular loops (prior to the third and fifth transmembrane domains) of most GPCRs (cys 121 and cys

197 of SEQ ID NO:2). A highly conserved asparagine residue in the first transmembrane domain is present (asn 67 in SEQ ID NO:2). Transmembrane domain two of the 52871 protein contains a highly conserved leucine (leu90 of SEQ ID NO:2). The two cysteine residues are believed to form a disulfide bond that stabilizes the functional protein structure. A highly conserved tryptophan and proline in the fourth transmembrane domain of the 52871 proteins is present (trp171 and pro 180 of SEQ ID NO:2). The third cytoplasmic loop contains 40 amino acid residues and is thus the longest cytoplasmic loop of the three, characteristic of G protein coupled receptors. Moreover, a highly conserved proline in the sixth transmembrane domain is present (pro 289 of SEQ ID NO:2). The proline residues in the fourth, fifth, sixth, and seventh transmembrane domains are thought to introduce kinks in the alpha-helices and may be important in the formation of the ligand binding pocket. Moreover, an almost invariant proline is present in the seventh transmembrane domain of 52871 (pro 327 of SEQ ID NO:2).

15 A search of the amino acid sequence of 52871 was also performed against the HMM database). This search resulted in the identification of a "7-TMR profile" ("7tm\_1" domain) in the amino acid sequence of 52871 (SEQ ID NO:2) at about residues 66-330 (score: 164.0). The results of this search are set forth in Figure 3.

Further domain motifs were identified by using the amino acid sequence of 20 52871 (SEQ ID NO:2) to search the ProDom database (<http://protein.toulouse.inra.fr/prodom.html>). Numerous matches against protein domains described as G-protein transmembrane domains and the like were identified.

A search was also performed against the Prosite database, and resulted in the identification of potential N-glycosylation sites at about residues 4-7 and 250-253, 25 (Prosite accession number PS0001). This search also identified the presence of a potential glycosaminoglycan attachment site (Prosite accession number PS0002) at about residues 13-16, two potential cAMP- and cGMP-dependant protein kinase phosphorylation sites at about residues 78-81 and 240-243 (Prosite accession number PS0004), two potential protein kinase C phosphorylation sites at about residues 47-49 30 and 74-76 (Prosite accession number PS0005), three potential casein kinase II phosphorylation sites at about residues 31-34, 115-118, and 191-194 (Prosite accession number PS0006), and two potential N-myristoylation sites at about residues 8-13, and 14-19 (Prosite accession number PS0008). Furthermore, this search also identified the presence of a potential amidation site at residues 185-188 (Prosite accession number 35 PS0009). This search also identified the presence of a G-protein coupled receptor signature motif at residues 134-150 (Prosite accession number PS00237). A structural, hydrophobicity, and antigenicity analysis of the human Fbh52871 protein was undertaken. The results of this analysis are set forth in Figure 2.

### Tissue Distribution of Human 52871 mRNA by Northern Analysis

This example describes the tissue distribution of 52871 mRNA, as determined by Northern analysis.

- 5 Northern blot hybridizations with the various RNA samples are performed under standard conditions and washed under stringent conditions, *i.e.*, 0.2xSSC at 65°C. The DNA probe is radioactively labeled with <sup>32</sup>P-dCTP using the Prime-It kit (Stratagene, La Jolla, CA) according to the instructions of the supplier. Filters containing human mRNA (MultiTissue Northern I and MultiTissue Northern II from Clontech, Palo Alto, CA) are probed in ExpressHyb hybridization solution (Clontech) and washed at high stringency according to manufacturer's recommendations.

### Tissue Distribution of Human 52871 mRNA by *In situ* Analysis

- 15 For *in situ* analysis, various tissues, *e.g.* tissues obtained from brain, spinal cord and skin from human, monkey, and rat, were first frozen on dry ice. Ten-micrometer-thick sections of the tissues were post-fixed with 4% formaldehyde in DEPC treated 1X phosphate-buffered saline at room temperature for 10 minutes before being rinsed twice in DEPC 1X phosphate-buffered saline and once in 0.1 M triethanolamine-HCl (pH 8.0). Following incubation in 0.25% acetic anhydride-0.1 M triethanolamine-HCl for 10 minutes, sections were rinsed in DEPC 2X SSC (1X SSC is 0.15M NaCl plus 0.015M sodium citrate). Tissue was then dehydrated through a series of ethanol washes, incubated in 100% chloroform for 5 minutes, and then rinsed in 100% ethanol for 1 minute and 95% ethanol for 1 minute and allowed to air dry.

- 25 Hybridizations were performed with <sup>35</sup>S-radiolabeled (5 X 10<sup>7</sup> cpm/ml) cRNA probes. Probes were incubated in the presence of a solution containing 600 mM NaCl, 10 mM Tris (pH 7.5), 1 mM EDTA, 0.01% sheared salmon sperm DNA, 0.01% yeast tRNA, 0.05% yeast total RNA type X1, 1X Denhardt's solution, 50% formamide, 10% dextran sulfate, 100 mM dithiothreitol, 0.1% sodium dodecyl sulfate (SDS), and 0.1% sodium thiosulfate for 18 hours at 55°C.

- 30 After hybridization, slides were washed with 2X SSC. Sections are then sequentially incubated at 37°C in TNE (a solution containing 10 mM Tris-HCl (pH 7.6), 500 mM NaCl, and 1 mM EDTA), for 10 minutes, in TNE with 10µg of RNase A per ml for 30 minutes, and finally in TNE for 10 minutes. Slides were then rinsed with 2X SSC at room temperature, washed with 2X SSC at 50°C for 1 hour, washed with 0.2X SSC at 55°C for 1 hour, and 0.2X SSC at 60°C for 1 hour. Sections were then dehydrated rapidly through serial ethanol-0.3 M sodium acetate concentrations before being air dried and exposed to Kodak Biomax MR scientific imaging film for 24 hours

and subsequently dipped in NB-2 photoemulsion and exposed at 4°C for 7 days before being developed and counter stained.

Results showed 52871 mRNA expression in rat brain (cortex and hippocampus), spinal cord, and dorsal root ganglia neurons, and testis. Results also showed 52871 mRNA expression in monkey cortex, dorsal root ganglia neurons, spinal cord, and testis. 52871 mRNA expression was also shown in human brain, spinal cord, DRG, and skin. *In situ* hybridization in monkey and rat tissues was performed with human 52871 probes. This cross-reactivity indicates that 52871 orthologues are likely highly conserved.

*In situ* hybridization in an animal model of pain, including models of pain caused by axotomized DRG, chronic constriction injury (CCI), and intraperitoneal administration of complete Freund's adjuvant (CFA), showed no regulation of 52871 mRNA expression, indicating that modulation of nociception by 52871 likely correlates with changes in 52871 activity as compared to changes in 52871 nucleic acid expression.

#### Tissue Expression Analysis of Human 52871 mRNA Using Taqman Analysis

This example describes the tissue distribution of human 52871 mRNA in a variety of cells and tissues, as determined using the TaqMan™ procedure. The Taqman™ procedure is a quantitative, reverse transcription PCR-based approach for detecting mRNA. The RT-PCR reaction exploits the 5' nuclease activity of AmpliTaq Gold™ DNA Polymerase to cleave a TaqMan™ probe during PCR. Briefly, cDNA was generated from the samples of interest, e.g., brain, testis, spinal cord, skin, dorsal root ganglia, placenta, etc., and used as the starting material for PCR amplification. In addition to the 5' and 3' gene-specific primers, a gene-specific oligonucleotide probe (complementary to the region being amplified) was included in the reaction (*i.e.*, the Taqman™ probe). The TaqMan™ probe includes the oligonucleotide with a fluorescent reporter dye covalently linked to the 5' end of the probe (such as FAM (6-carboxyfluorescein), TET (6-carboxy-4,7,2',7'-tetrachlorofluorescein), JOE (6-carboxy-4,5-dichloro-2,7-dimethoxyfluorescein), or VIC) and a quencher dye (TAMRA (6-carboxy-N,N,N',N'-tetramethylrhodamine) at the 3' end of the probe.

During the PCR reaction, cleavage of the probe separates the reporter dye and the quencher dye, resulting in increased fluorescence of the reporter. Accumulation of PCR products is detected directly by monitoring the increase in fluorescence of the reporter dye. When the probe is intact, the proximity of the reporter dye to the quencher dye results in suppression of the reporter fluorescence. During PCR, if the target of interest is present, the probe specifically anneals between the forward and reverse primer sites. The 5'-3' nucleolytic activity of the AmpliTaq™ Gold DNA Polymerase cleaves

the probe between the reporter and the quencher only if the probe hybridizes to the target. The probe fragments are then displaced from the target, and polymerization of the strand continues. The 3' end of the probe is blocked to prevent extension of the probe during PCR. This process occurs in every cycle and does not interfere with the exponential accumulation of product. RNA was prepared using the trizol method and treated with DNase to remove contaminating genomic DNA. cDNA was synthesized using standard techniques. Mock cDNA synthesis in the absence of reverse transcriptase resulted in samples with no detectable PCR amplification of the control gene confirms efficient removal of genomic DNA contamination.

A human normal tissue panel indicated that human 52871 is expressed at very low levels. The highest expression was in human brain, followed by spinal cord and dorsal root ganglia (DRG) (*see* Table 1, below).

**TABLE 1. Tissue Expression Analysis of Human 52871 mRNA Using Taqman Analysis**

<u>Tissue</u>	<u>52871</u>	<u>Beta 2</u> <u>Mean</u>	<u>Expression</u>
Adrenal Gland	38.77	20.17	0.00
Brain	33.02	20.42	0.28
Heart	39.46	20.42	0.00
Kidney	40.00	19.46	0.00
Liver	40.00	20.14	0.00
Lung	40.00	18.37	0.00
Mammary Gland	40.00	19.23	0.00
Placenta	36.91	20.11	0.02
Prostate	39.46	19.41	0.00
Salivary Gland	40.00	20.53	0.00
Muscle	40.00	22.24	0.01
Small Intestine	40.00	19.52	0.00
Spleen	39.12	17.99	0.00
Stomach	40.00	19.83	0.00
Testes	32.84	21.36	0.61
Thymus	38.63	19.57	0.00
Trachea	40.00	20.77	0.00
Uterus	40.00	20.49	0.00
Spinal Cord	34.24	20.79	0.16
Skin	37.03	18.68	0.01
DRG	35.11	20.51	0.07

**EXAMPLE 2:            EXPRESSION OF RECOMBINANT 52871 PROTEIN IN  
BACTERIAL CELLS**

5            In this example, 52871 is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, 52871 is fused to GST and this fusion polypeptide is expressed in *E. coli*, e.g., strain PEB199. Expression of the GST-52871 fusion protein in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from  
10 crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

15

**EXAMPLE 3:            EXPRESSION OF RECOMBINANT 52871 PROTEIN  
IN COS CELLS**

          To express the 52871 gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of  
20 replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire 52871 protein and an HA tag (Wilson *et al.* (1984) *Cell* 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant  
25 protein under the control of the CMV promoter.

          To construct the plasmid, the 52871 DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the 52871 coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other  
30 restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the 52871 coding sequence. The PCR amplified fragment and the pcDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the 52871 gene is  
35 inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5 $\alpha$ , SURE, available from Stratagene Cloning Systems, La Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and

resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

COS cells are subsequently transfected with the 52871-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the 52871 polypeptide is detected by radiolabelling ( $^{35}\text{S}$ -methionine or  $^{35}\text{S}$ -cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988) using an HA specific monoclonal antibody. Briefly, the cells are labeled for 8 hours with  $^{35}\text{S}$ -methionine (or  $^{35}\text{S}$ -cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA-specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

Alternatively, DNA containing the 52871 coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the 52871 polypeptide is detected by radiolabelling and immunoprecipitation using a 52871 specific monoclonal antibody.

### Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.



What is claimed:

1. An isolated nucleic acid molecule selected from the group consisting of:
  - a) a nucleic acid molecule comprising the nucleotide sequence set forth in  
5 SEQ ID NO:1; and
  - b) a nucleic acid molecule comprising the nucleotide sequence set forth  
in SEQ ID NO:3.
2. An isolated nucleic acid molecule which encodes a polypeptide  
10 comprising the amino acid sequence set forth in SEQ ID NO:2.
3. An isolated nucleic acid molecule comprising the nucleotide sequence  
contained in the plasmid deposited with ATCC® as Accession Number \_\_\_\_\_.
- 15 4. An isolated nucleic acid molecule which encodes a naturally occurring  
allelic variant of a polypeptide comprising the amino acid sequence set forth in SEQ ID  
NO:2.
5. An isolated nucleic acid molecule selected from the group consisting of:
  - 20 a) a nucleic acid molecule comprising a nucleotide sequence which is at  
least 60% identical to the nucleotide sequence of SEQ ID NO:1 or 3, or a complement  
thereof;
  - b) a nucleic acid molecule comprising a fragment of at least 30  
nucleotides of a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1 or 3,  
25 or a complement thereof;
  - c) a nucleic acid molecule which encodes a polypeptide comprising an  
amino acid sequence at least about 60% identical to the amino acid sequence of SEQ ID  
NO:2; and
  - d) a nucleic acid molecule which encodes a fragment of a polypeptide  
30 comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises  
at least 10 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:2.
6. An isolated nucleic acid molecule which hybridizes to a complement of  
the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 under stringent conditions.  
35
7. An isolated nucleic acid molecule comprising a nucleotide sequence  
which is complementary to the nucleotide sequence of the nucleic acid molecule of any  
one of claims 1, 2, 3, 4, or 5.

8. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5, and a nucleotide sequence encoding a heterologous polypeptide.
- 5
9. A vector comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.
10. The vector of claim 9, which is an expression vector.
- 10
11. A host cell transfected with the expression vector of claim 10.
12. A method of producing a polypeptide comprising culturing the host cell of claim 11 in an appropriate culture medium to, thereby, produce the polypeptide.
- 15
13. An isolated polypeptide selected from the group consisting of:
- a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 10 contiguous amino acids of SEQ ID NO:2;
- 20
- b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a complement of a nucleic acid molecule consisting of SEQ ID NO:1 or 3 under stringent conditions;
- c) a polypeptide which is encoded by a nucleic acid molecule comprising
- 25 a nucleotide sequence which is at least 60% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1 or 3; and
- d) a polypeptide comprising an amino acid sequence which is at least 60% identical to the amino acid sequence of SEQ ID NO:2.
- 30
14. The isolated polypeptide of claim 13 comprising the amino acid sequence of SEQ ID NO:2.
15. The polypeptide of claim 13, further comprising heterologous amino acid sequences.
- 35
16. An antibody which selectively binds to a polypeptide of claim 13.

17. A method for detecting the presence of a polypeptide of claim 13 in a sample comprising:

a) contacting the sample with a compound which selectively binds to the polypeptide; and

5 b) determining whether the compound binds to the polypeptide in the sample to thereby detect the presence of a polypeptide of claim 13 in the sample.

18. The method of claim 17, wherein the compound which binds to the polypeptide is an antibody.

10

19. A kit comprising a compound which selectively binds to a polypeptide of claim 13 and instructions for use.

20. A method for detecting the presence of a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in a sample comprising:

15

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to a complement of the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to the complement of the nucleic acid molecule in the sample to thereby detect the presence of  
20 the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in the sample.

21. The method of claim 20, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

22. A kit comprising a compound which selectively hybridizes to a complement of the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 and instructions for use.

25

23. A method for identifying a compound which binds to a polypeptide of claim 13 comprising:

30

a) contacting the polypeptide, or a cell expressing the polypeptide with a test compound; and

b) determining whether the polypeptide binds to the test compound.

24. The method of claim 23, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detection of test compound/polypeptide binding;

5 b) detection of binding using a competition binding assay; and

c) detection of binding using an assay for 52871 activity.

25. A method for modulating the activity of a polypeptide of claim 13 comprising contacting the polypeptide or a cell expressing the polypeptide with a  
10 compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

26. A method for identifying a compound which modulates the activity of a polypeptide of claim 13 comprising:

15 a) contacting the polypeptide, or a cell expressing the polypeptide with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

20

27. The method of claim 26, wherein said activity is modulation of nociception.

28. A method for identifying a compound which modulates pain comprising:

25 a) contacting the polypeptide of claim 13, or a cell expressing the polypeptide with a test compound; and

b) identifying the compound as a modulator of pain by determining the effect of the test compound on the activity of the polypeptide.

30 29. A method for identifying a compound which modulates nociception comprising:

a) contacting the polypeptide of claim 13, or a cell expressing the polypeptide with a test compound; and

35 b) identifying the compound as a modulator of nociception by determining the effect of the test compound on the activity of the polypeptide.

30. A method for treating a subject having pain or a pain disorder comprising administering to the subject a 52871 modulator, thereby treating said subject having a pain disorder.

5 31. A method for treating a subject having pain or a pain disorder comprising administering to the subject a 52871 modulator, wherein the 52871 modulator is the modulator identified by the method of claim 26, thereby treating said subject having a pain disorder.

10 32. The method of claim 30, wherein the 52871 modulator is a small molecule.

33. The method of claim 30, wherein said 52871 modulator is administered in a pharmaceutically acceptable formulation.

15

34. The method of claim 30, wherein said 52871 modulator is administered using a gene therapy vector.

35. The method of 30, wherein the 52871 modulator is capable of modulating  
20 52871 polypeptide activity.

Input file Fbh52871FL.seq; Output File 52871.trans  
Sequence length 1731

AATTCCCGAGTCGACCCACGCGTCCGAGCACGTAGATCCTCCCTGTCTCATCAGGCAGAGCTCTTCAGTGAGGTGGGCTCA  
GGGAGGGCTCTGTGCCCTCTGTTTCAGCAGAGCTGCAGCTGCTGCCCAGCTCTCAGGAGGCAAGCTGGACTCCCTCACTCG  
GCTGCAGGAGCAAGGACAGTGAGGCTCAACCCCGCCTGAGCC ATG CCA GCC AAC TTC ACA GAG GGC AGC 9  
27  
F D S S G T G Q T L D S S P V A C T E A 29  
TTC GAT TCC AGT GGG ACC GGG CAG ACG CTG GAT TCT TCC CCA GTG GCT TGC ACT GAA GCA 87  
V T F T E V V K G K E W G S F Y Y S F K 49  
GTG ACT TTT ACT GAA GTG GTG AAA GGA AAG GAA TGG GGT TCC TTC TAC TAC TCC TTT AAG 147  
T E Q L I T L W V L F V F T I V G N S V 69  
ACT GAG CAA TTG ATA ACT CTG TGG GTC CTC TTT GTT TTT ACC ATT GTT GGA AAC TCC GTT 207  
V L F S T W R R K K K S R M T F F V T Q 89  
GTG CTT TTT TCC ACA TGG AGG AGA AAG AAG AAG TCA AGA ATG ACC TTC TTT GTG ACT CAG 267  
L A I T D S F T G L V N I L T D I I W R 109  
CTG GCC ATC ACA GAT TCT TTC ACA GGA CTG GTC AAC ATC TTG ACA GAT ATT ATT TGG CGA 327  
F T G D F T A P D L V C R V V R Y L Q V 129  
TTC ACC GGA GAC TTC ACG GCA CCT GAC CTG GTT TGC CGA GTG GTC CGC TAT TTG CAG GTT 387  
V L L Y A S T Y V L V S L S I D R Y H A 149  
GTG CTG CTG TAC GCC TCT ACC TAC GTC CTG GTG TCC CTC AGC ATA GAC AGA TAC CAT GCC 447  
I V Y P M K F L Q G E K Q A R V L I V I 169  
ATC GTC TAC CCC ATG AAG TTC CTT CAA GGA GAA AAG CAA GCC AGG GTC CTC ATT GTG ATC 507  
A W S L S F L F S I P T L I I F G K R T 189  
GCC TGG AGC CTG TCT TTT CTG TTC TCC ATT CCC ACC CTG ATC ATA TTT GGG AAG AGG ACA 567  
L S N G E V Q C W A L W P D D S Y W T P 209  
CTG TCC AAC GGT GAA GTG CAG TGC TGG GCC CTG TGG CCT GAC GAC TCC TAC TGG ACC CCA 627  
Y M T I V A F L V Y F I P L T I I S I M 229  
TAC ATG ACC ATC GTG GCC TTC CTG GTG TAC TTC ATC CCT CTG ACA ATC ATC AGC ATC ATG 687  
Y G I V I R T I W I K R K T Y E T V I S 249  
TAT GGC ATT GTG ATC CGA ACT ATT TGG ATT AAA AGG AAA ACC TAC GAA ACA GTG ATT TCC 747  
N C S D G K L C S S Y N R G L I S K A K 269  
AAC TGC TCA GAT GGG AAA CTG TGC AGC AGC TAT AAC CGA GGA CTC ATC TCA AAG GCA AAA 807  
I K A I K Y S I I I I L A F I C C W S P 289  
ATC AAG GCT ATC AAG TAT AGC ATC ATC ATC ATT CTT GCC TTC ATC TGC TGT TGG AGT CCA 867  
Y F L F D I L D N F N L L P D T Q E R F 309  
TAC TTC CTG TTT GAC ATT TTG GAC AAT TTC AAC CTC CTT CCA GAC ACC CAG GAG CGT TTC 927  
Y A S V I I Q N L P A L N S A I N P L I 329  
TAT GCC TCT GTG ATC ATT CAG AAC CTG CCA GCA TTG AAT AGT GCC ATC AAC CCC CTC ATC 987  
Y C V F S S S I S F P C R I I D G N D \* 349  
TAC TGT GTC TTC AGC AGC TCC ATC TCT TTC CCC TGC AGG ATC ATA GAT GGA AAT GAC TAG 1047  
CCTGTCTCAGATGACACTTCGAACCTTTGGACTTTTGAGTTAATGTTGGAATAAGTTAAGACTTTTCGGGGACTGTTGGG  
AAGGCAGGATTGTATTTTGAAATTTGAGAAGGACATAAAATTTGGGAGGGGGCAGCATGGAATCATATGGTCTAGATAT

Figure 1A

ATGACCCTGTCCAAATCTCAAATCTAACTGTAATTCACAGTGTTGGAGGTGGGGTCTGGTGGGAGGTGATTGATCATG  
GAGGTGGAGTTCTCATTAATGATTTAGAGCCATCCCTTTTGTTATGGTATAGTGAGTGAGTTATCACAAGATCTCTTTG  
TTTAAAAGTTGTGGTACCTCCACCTCTCTCTCTTGCTCCTGCTCTGGCCATGTAAGACGTGCCTGCTTCCCCTTCAC  
CTTCTTGCAATGATTGTAAGTTTCCTGAGGCCTCCCCAAAAGCAGAAGCCACTATGCTTCCTGAACAGCCAATGGAACCT  
CGTGCCAAAA

**Figure 1B**

## Analysis of 52871 (348 aa)

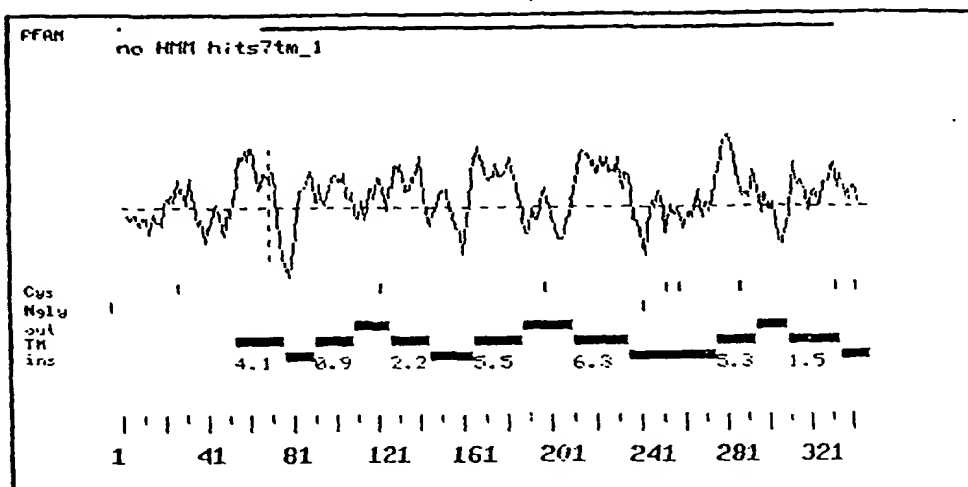


Figure 2



## Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM

hmmpfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam5.3/Pfam  
Sequence file: /prod/ddm/wspace/orfanal/oa-script.24854.seq

Query: 52871

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value
7tm_1	7 transmembrane receptor (rhodopsin family)	164.0	6.1e-51

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
7tm_1	1/1	66	330	1	259	164.0	6.1e-51

Alignments of top-scoring domains:

7tm\_1: domain 1 of 1, from 66 to 330: score 164.0, E = 6.1e-51

```

*->GNLLVilvilrtkkklrtptnifilNLAVADLLflltlppwalyylvg
52871 66 GN++V++ +r +k ++ + +f+ LA++D +l++ + +++ +
GNSVVLFFSTWR-RKKKSRMTFFVTQLAITDSFTGLVNILTDIIWRFT 111

gsedWpfGsalCklvtaldvvnmyaSillLtaISiDRYlAivhPlryrrr
52871 112 g d ++++C++v +l+vv +yaS + L+ +SiDRY Aiv+P++ +
G--DFTAPDLVCRVVRVYLQVLLYASTYVLVSLSIDRYHAIVYPMKFLQG 159

rtsprrrAkvvillvWvlalllslPpllfswwktveegnngtlnvntvCli
52871 160 A+v+i+++W l++l+s+P l+ ++ +t ++g +C+
EK---QARVLIVIAWSLSFLFSIPTLIIFGKRTLSNG-----EVQCWA 199

dfpeestasvstwlrsyvllstlvGFl1Pl1vilvcYtrIlrtlr.....
52871 200 +p +s ++ +y ++ +++++Pl++i+++Y+ ++rt++ ++++
LWPDDS-----YWTPYMTIVAFLVYFIPLTIISIMYGIVIRTIWIKRKT 243

.....kaaktllvvvvvFv1CWlPyfiv
52871 244 ++ ++ ++++ ++ +++ ++ + ka k +++++F++CW Pyf++
yetvisncsdgklcssynrgliskakiKAIKYSIIIIILAFICCWSPYFLF 293

llldtlc.lsiimsstCelervlptallvtlwLayvNsc1NPiIY<--
52871 294 +ld ++ l+ + +a ++ L +Ns++NP+IY
DILDNFN1LP-----DTQERFYASVIIQNLPALNSAINPLIY 330
  
```

Figure 3

## SEQUENCE LISTING

&lt;110&gt; Millennium Pharmaceuticals, Inc., et al.

<120> 52871, A NOVEL HUMAN G PROTEIN COUPLED  
RECEPTOR AND USES THEREOF

&lt;130&gt; MNI-165PC

&lt;150&gt; USSN 60/212,331

&lt;151&gt; 2000-06-16

&lt;150&gt; USSN 60/269,758

&lt;151&gt; 2001-02-16

&lt;160&gt; 8

&lt;170&gt; FastSEQ for Windows Version 4.0

&lt;210&gt; 1

&lt;211&gt; 1731

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (201)...(1244)

&lt;400&gt; 1

```

aattcccgag tgcacccacg cgtccgagca cgtagatcct ccctgtcatc aggcagagct 60
cttcagttag gtgggctcag ggagggctct gtgcctctgt tcagcagagc tgcagctgct 120
gcccagctct caggaggcaa gctggactcc ctactcggc tgcaggagca aggacagtga 180
ggctcaaccc cgcctgagcc atg cca gcc aac ttc aca gag gcc agc ttc gat 233
                Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp
                  1             5             10

```

```

tcc agt ggg acc ggg cag acg ctg gat tct tcc cca gtg gct tgc act 281
Ser Ser Gly Thr Gly Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr
          15             20             25

```

```

gaa gca gtg act ttt act gaa gtg gtg aaa gga aag gaa tgg ggt tcc 329
Glu Ala Val Thr Phe Thr Glu Val Val Lys Gly Lys Glu Trp Gly Ser
          30             35             40

```

```

ttc tac tac tcc ttt aag act gag caa ttg ata act ctg tgg gtc ctc 377
Phe Tyr Tyr Ser Phe Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu
          45             50             55

```

```

ttt gtt ttt acc att gtt gga aac tcc gtt gtg ctt ttt tcc aca tgg 425
Phe Val Phe Thr Ile Val Gly Asn Ser Val Val Leu Phe Ser Thr Trp
          60             65             70             75

```

```

agg aga aag aag aag tca aga atg acc ttc ttt gtg act cag ctg gcc 473
Arg Arg Lys Lys Lys Ser Arg Met Thr Phe Phe Val Thr Gln Leu Ala
          80             85             90

```

```

atc aca gat tct ttc aca gga ctg gtc aac atc ttg aca gat att att 521
Ile Thr Asp Ser Phe Thr Gly Leu Val Asn Ile Leu Thr Asp Ile Ile
          95             100             105

```

```

tgg cga ttc acc gga gac ttc acg gca cct gac ctg gtt tgc cga gtg 569

```

Arg	Val	Leu	Leu	Leu	Ile	Ser	Gly	Val	Trp	Leu	Leu	Ser	Leu	Leu	Ile
225					230					235					240
Ser	Ser	Pro	Pro	Leu	Ile	Gly	Trp	Asn	Asp	Trp	Pro	Asp	Glu	Phe	Thr
				245					250					255	
Ser	Ala	Thr	Pro	Cys	Glu	Leu	Thr	Ser	Gln	Arg	Gly	Tyr	Val	Ile	Tyr
			260					265					270		
Ser	Ser	Leu	Gly	Ser	Phe	Phe	Ile	Pro	Leu	Ala	Ile	Met	Thr	Ile	Val
		275					280					285			
Tyr	Ile	Glu	Ile	Phe	Val	Ala	Thr	Arg	Arg	Arg	Leu	Arg	Glu	Arg	Ala
	290					295					300				
Arg	Ala	Asn	Lys	Leu	Asn	Thr	Ile	Ala	Leu	Lys	Ser	Thr	Glu	Leu	Glu
305					310					315					320
Pro	Met	Ala	Asn	Ser	Ser	Pro	Val	Ala	Ala	Ser	Asn	Ser	Gly	Ser	Lys
				325					330					335	
Ser	Arg	Leu	Leu	Ala	Ser	Trp	Leu	Cys	Cys	Gly	Arg	Asp	Arg	Ala	Gln
		340						345					350		
Phe	Ala	Thr	Pro	Met	Ile	Gln	Asn	Asp	Gln	Glu	Ser	Ile	Ser	Ser	Glu
		355					360					365			
Thr	His	Gln	Pro	Gln	Asp	Ser	Ser	Lys	Ala	Gly	Pro	His	Gly	Asn	Ser
	370					375					380				
Asp	Pro	Gln	Gln	Gln	His	Val	Val	Val	Leu	Val	Lys	Lys	Ser	Arg	Arg
385					390					395					400
Ala	Lys	Thr	Lys	Asp	Ser	Ile	Lys	His	Gly	Lys	Thr	Arg	Gly	Gly	Arg
				405					410					415	
Lys	Ser	Gln	Ser	Ser	Ser	Thr	Cys	Glu	Pro	His	Gly	Glu	Gln	Gln	Leu
		420						425					430		
Leu	Pro	Ala	Gly	Gly	Asp	Gly	Gly	Ser	Cys	Gln	Pro	Gly	Gly	Gly	His
		435					440					445			
Ser	Gly	Gly	Gly	Lys	Ser	Asp	Ala	Glu	Ile	Ser	Thr	Glu	Ser	Gly	Ser
	450					455					460				
Asp	Pro	Lys	Gly	Cys	Ile	Gln	Val	Cys	Val	Thr	Gln	Ala	Asp	Glu	Gln
465					470					475					480
Thr	Ser	Leu	Lys	Leu	Thr	Pro	Pro	Gln	Ser	Ser	Thr	Gly	Val	Ala	Ala
				485					490					495	
Val	Ser	Val	Thr	Pro	Leu	Gln	Lys	Lys	Thr	Ser	Gly	Val	Asn	Gln	Phe
		500						505					510		
Ile	Glu	Glu	Lys	Gln	Lys	Ile	Ser	Leu	Ser	Lys	Glu	Arg	Arg	Ala	Ala
	515						520					525			
Arg	Thr	Leu	Gly	Ile	Ile	Met	Gly	Val	Phe	Val	Ile	Cys	Trp	Leu	Pro
	530					535					540				
Phe	Phe	Leu	Met	Tyr	Val	Ile	Leu	Pro	Phe	Cys	Gln	Thr	Cys	Cys	Pro
545					550					555					560
Thr	Asn	Lys	Phe	Lys	Asn	Phe	Ile	Thr	Trp	Leu	Gly	Tyr	Ile	Asn	Ser
				565					570					575	
Gly	Leu	Asn	Pro	Val	Ile	Tyr	Thr	Ile	Phe	Asn	Leu	Asp	Tyr	Arg	Arg
			580					585					590		
Ala	Phe	Lys	Arg	Leu	Leu	Gly	Leu	Asn							
		595					600								

Applicant's or agent's file reference  <b>MNI-165PC</b>	International Application No.  <b>PCT/US01/19464</b>
--	--

**INDICATIONS RELATING TO DEPOSITED MICRO-ORGANISM  
OR OTHER BIOLOGICAL MATERIAL**

(PCT Rule 13bis)

<p><b>A.</b> The indications made below relate to the deposited micro-organisms or other biological material referred to in the description at the instance "_____" on: page 2, line 19; page 12, line 25; page 14, line 1; page 14, lines 37; page 15, line 3; page 15, line 10; page 15, line 13; page 15, line 35; page 15, line 38; page 16, line 3; page 16, line 5; page 16, line 13; page 16, line 20; page 16, line 23; page 16, line 35; page 16, line 36; page 16, line 36; page 17, line 2; page 17, line 27; page 17, line 36; page 18, line 2; page 18, line 7; page 18, line 28; page 19, line 17; page 19, line 22; page 20, line 1; page 21, line 16; page 21, line 21; page 22, line 5; page 22, line 8; page 22, line 26; page 25, line 3; page 28, line 32; page 28, line 36; page 43, line 10; page 67, line 2, and the second "_____" on page 85, line 21.</p>	
<p><b>B. IDENTIFICATION OF DEPOSIT</b> Further deposits are identified on an additional sheet. <input type="checkbox"/></p>	
<p>Name of depositary institution  <b>American Type Culture Collection (ATCC)</b> Address of depositary institution (including postal code and country)  <b>10801 University Boulevard, Manassas, VA 20110-2209 USA</b></p>	
Date of deposit  <b>01 June 2001</b>	Accession Number  <b>PTA-3426</b>
<p><b>C. ADDITIONAL INDICATIONS</b> (leave blank if not applicable) This information is continued on an additional sheet. <input type="checkbox"/></p>	
<p><b>D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE</b> (if the indications are not for all designated States)</p>	
<p><b>E. SEPARATE FURNISHING OF INDICATIONS</b> (leave blank if not applicable)</p>	
<p>The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")</p>	

<p><b>For Receiving Office use only</b></p> <p>This sheet was received with the International Application</p> <p>Authorized Officer</p>	<p><b>For International Bureau use only</b></p> <p>This sheet was received by the International Bureau on:</p> <p align="center"><b>17. 11. 01</b></p> <p>Authorized Officer</p>
---	--

Form PCT/RO/134 (July 1998)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
20 December 2001 (20.12.2001)

PCT

(10) International Publication Number  
**WO 01/96400 A3**

(51) International Patent Classification<sup>7</sup>: C07K 14/72,  
G01N 33/566

(21) International Application Number: PCT/US01/19464

(22) International Filing Date: 18 June 2001 (18.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/212,331 16 June 2000 (16.06.2000) US  
60/269,758 16 February 2001 (16.02.2001) US

(71) Applicant (for all designated States except US): MIL-  
LENNIUM PHARMACEUTICALS, INC. [US/US]; 75  
Sidney Street, Cambridge, MA 02139 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): GLUCKSMANN,  
Maria, Alexandra [AR/US]; 33 Summit Road, Lexing-  
ton, MA 02173 (US). SILOS-SANTIAGO, Inmaculada  
[ES/US]; 18 Hillard Street, Cambridge, MA 02138 (US).

(74) Agents: MANDRAGOURAS, Amy, E.; Lahive & Cock-  
field, LLP, 28 State Street, Boston, MA 02109 et al. (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,

CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,  
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,  
ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**

- with international search report
- before the expiration of the time limit for amending the  
claims and to be republished in the event of receipt of  
amendments
- with (an) indication(s) in relation to deposited biological  
material furnished under Rule 13bis separately from the  
description

(88) Date of publication of the international search report:  
25 April 2002

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: 52871, A NOVEL HUMAN G PROTEIN COUPLED RECEPTOR AND USES THEREOF

(57) Abstract: The invention provides isolated nucleic acids molecules, designated 52871 nucleic acid molecules, which encode novel G-protein Coupled Receptor molecules. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing 52871 nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which a 52871 gene has been introduced or disrupted. The invention still further provides isolated 52871 proteins, fusion proteins, antigenic peptides and anti-52871 antibodies. Diagnostic methods utilizing compositions of the invention are also provided.

WO 01/96400 A3

## INTERNATIONAL SEARCH REPORT

In tional Application No

PCT/US 01/19464

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 C07K14/72 G01N33/566

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C07K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, MEDLINE

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	WO 01 18206 A (LEXICON GENETICS INC) 15 March 2001 (2001-03-15)  98,842% identity of SEQ-ID1 in a 1036nt overlap with present SEQ-ID1; almost 100% identity of SEQ-ID2 with present SEQ-ID2 ---	4-13, 15-18, 20-35
E	WO 01 48015 A (UPJOHN CO) 5 July 2001 (2001-07-05)  98,914% identity of SEQ-ID12 in a 1105nt overlap with present SEQ-ID1; almost 100% identity of SEQ-ID25 with present SEQ-ID2 --- -/--	4-13, 15-18, 20-35



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

## \* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \* & \* document member of the same patent family

Date of the actual completion of the international search

31 January 2002

Date of mailing of the international search report

26/02/2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Goetz, M

## INTERNATIONAL SEARCH REPORT

 International Application No  
 PCT/US 01/19464

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EM_HTG 'Online! EMBL; Accession-n° AC021182, 16 January 2000 (2000-01-16) WATERSTONE R.H.: "The sequence of Homo Sapiens Clone" XP002188802	5-8
Y	Homo Sapiens chromosome 7 clone; 99,065% identity in 321nt overlap with present SEQ-ID1; cited against option b) of present claim 5 abstract	9-12, 20-22
X	--- DATABASE EM_HUM 'Online! EMBL; Accession-n° AC018648, 21 December 1999 (1999-12-21) KAUL R.K. ET AL.: "Large-scale mapping and sequencing of human chromosome 7" XP002188803	5-8
Y	Homo Sapiens chromosome 7 clone; 99,979% identity in 521nt overlap with present SEQ-ID1; cited against option b) of present claim 5 abstract	9-12, 20-22
X	--- DATABASE EM_HTG 'Online! EMBL; Accession-n° AC023002, 8 February 2000 (2000-02-08) BIRREN B. ET AL.: "Homo sapiens chromosome 7, clone RP11-265E18" XP002188804	5-8
Y	Homo Sapiens chromosome 7 clone; 87,586% identity in 435nt overlap with present SEQ-ID1; cited against option b) of present claim 5 abstract	9-12, 20-22
X	--- DATABASE EM_HUM 'Online! EMBL; Accession-n° AC005853, 23 October 1998 (1998-10-23) IADONATO S.P. ET AL.: "Large-scale mapping and sequencing of human chromosome 7" XP002188805	5-8
Y	Homo Sapiens clone UWGC:g1564; 97,674% identity in 215nt overlap with present SEQ-ID3; cited against option b) of present claim 5 abstract	9-12, 20-22
	--- -/--	

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 01/19464

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EM_EST 'Online! EMBL; Accession-n° HSA04400, 25 July 1996 (1996-07-25) HILLIER L. ET AL.: "Generation and analysis of 280,000 human ESTs" XP002188806	5-8
Y	zh91c08.r1 Soares_fetal_liver; 97,842% identity in 139nt overlap with present SEQ-ID3; cited against option b) of present claim 5 abstract & GENOME RESEARCH, vol. 6, no. 9, September 1996 (1996-09), pages 807-828, ---	9-12, 20-22
Y	WO 00 12707 A (MILLENNIUM PHARM INC) 9 March 2000 (2000-03-09) the whole document ---	1-35
Y	LEE N H ET AL: "MOLECULAR BIOLOGY OF G-PROTEIN-COUPLED RECEPTORS" DRUG NEWS AND PERSPECTIVES, XX, XX, vol. 6, no. 7, 1 September 1993 (1993-09-01), pages 488-497, XP000677175 ISSN: 0214-0934 the whole document ---	1-35
Y	KILPATRICK G J ET AL: "7TM receptors: the splicing on the cake" TRENDS IN PHARMACOLOGICAL SCIENCES, ELSEVIER TRENDS JOURNAL, CAMBRIDGE, GB, vol. 20, no. 7, 1 July 1999 (1999-07-01), pages 294-301, XP004171850 ISSN: 0165-6147 the whole document ---	1-35
Y	WILSON S ET AL: "ORPHAN G-PROTEIN-COUPLED RECEPTORS: THE NEXT GENERATION OF DRUG TARGETS?" BRITISH JOURNAL OF PHARMACOLOGY, BASINGSTOKE, HANTS, GB, vol. 125, no. 7, December 1998 (1998-12), pages 1387-1392, XP001010584 ISSN: 0007-1188 the whole document --- -/--	1-35



# INTERNATIONAL SEARCH REPORT

Ir. ational Application No

PCT/US 01/19464

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>STADEL J M ET AL: "Orphan G protein-coupled receptors: a neglected opportunity for pioneer drug discovery" TRENDS IN PHARMACOLOGICAL SCIENCES, ELSEVIER TRENDS JOURNAL, CAMBRIDGE, GB, vol. 18, no. 11, 1 November 1997 (1997-11-01), pages 430-437, XP004099345 ISSN: 0165-6147 the whole document -----</p>	1-35

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/US 01/19464

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
WO 0118206	A	15-03-2001	AU	7475400 A	10-04-2001
			WO	0118206 A1	15-03-2001
WO 0148015	A	05-07-2001	AU	5787501 A	09-07-2001
			WO	0148015 A2	05-07-2001
			AU	4165801 A	03-09-2001
			WO	0162797 A2	30-08-2001
WO 0012707	A	09-03-2000	AU	5801099 A	21-03-2000
			WO	0012707 A1	09-03-2000